TALLINN UNIVERSITY OF TECHNOLOGY

School of Business and Governance

Department of Business Administration

Anton Mikael Kallio

INVESTMENT RELATED INFORMATION GATHERING BY DIFFERENT AGE AND EDUCATION GROUPS: THE EFFECT OF OVERCONFIDENCE TO INFORMATION SOURCES

Bachelor's thesis

Programme - International Business Administration, Finance and Accounting

Supervisor: Pavlo Illiashenko

Tallinn 2021

I hereby declare that I have compiled the thesis independently

and all works, important standpoints and data by other authors

have been properly referenced and the same paper

has not been previously presented for grading.

The document length is 9017 words from the introduction to the end of conclusion.

Anton Kallio ……………………………

(signature, date)

Student code: 183941TVTB

Student e-mail address: anton.m.kallio@gmail.com

Supervisor: Pavlo Illiashenko:

The paper conforms to requirements in force

……………………………………………

(signature, date)

Chairman of the Defence Committee:

Permitted to the defence

…………………………………

(name, signature, date)

2

TABLE OF CONTENTS

ABSTRACT

This bachelor's thesis investigates the association between socio-demographic characteristics and overconfidence on the one hand and the likelihood of using specific source of information relevant for investing on the other. The three main sources are (1) word-of-mouth, social media, and other press; (2) banks, investing experts/analysts, and specialized press; and (3) company reports, and annual general meetings. Using the survey data from 140 individuals, the study runs a total of 18 logistic regressions to test the two main hypotheses. The results show that overconfidence had an influence on the used information sources, similarly as in the previous studies. Overconfidence has a negative association with using word-of-mouth, social media, and other press information sources. Age showed significance on some of the information sources, but not on all the predicted sources. Older people use word-of-mouth, social media, and other press less than younger people. Direct level of education was not found to be significant factor for the investor's choice of used sources of information. Experience and financial knowledge were found to have higher significance than the level of education. Experienced people are more likely to use company reports and annual general meetings as information sources, and less likely to use word-of mouth, social media, and other press. Based on this study some suggestions for future investigation of the subject are given.

Keywords: Sources of information, Overconfidence, Age, Education

INTRODUCTION

Investors are often relentless information gatherers, but there are also those who let other people do the work for them and use their knowledge. Some investors read and analyse companies' financial reports to make their investment decisions, while others trust that the analyst did their work well. Some investors do not really care about seeking information, so they scroll through social media and decide their investment assets by information given by others who use social media. Abreu and Mendes (2012) studied the effect of overconfidence and information sources to the frequency of trading and found that both aspects have an influence on the investor.

One of the main reasons for seeking information about investing is that the investor wants to make profitable decisions and to do better than the average market portfolio does. In the past 20 years the use and possibility to access the Internet has increased substantially. These days most people are able find the information they need within minutes or hours. This brings up the questions what sources they are using and is there demographic or behavioral factors that could make a difference? Several studies have concluded that people with higher education and financial literacy show better returns and ability to find good trades (Bucher-Koenen, Ziegelmeyer 2014; Talpsepp, Liivamägi, Vaarmets 2020). If education is a factor influencing investment returns, is it because they have been taught to use or trust certain information sources more? Young investors may not have been able to study at higher levels or receive great amount of knowledge about investing. They are still able to access a lot of the same information as the other investors are. Are the younger investors still finding information from different sources than the older investors because, for example, it is more convenient to use familiar social media platforms than lengthy company reports? Overconfidence has an influence on the investors trading habits (Abreu, Mendes 2012). Abreu and Mendes recognized that overconfident investors trade less when they receive information from friends and family. Non-overconfident traders have a higher probability to trade when they get their information from specialized sources. Overconfident investors may be prone to use different sources than those non-overconfident people with similar backgrounds because they exaggerate the quality of their own information (Xia *et al*. 2014).

My motivation to study this topic is to show that different factors could affect the information sources investors use. This knowledge could help analysts, company managers and all the investors to understand different occurrences in the stock market. Understanding the differences between investors information gathering could especially help young investors like myself evaluate if my investing strategy is viable. I also hope to increase amount of knowledge about the effect of age and education to investor's main information sources.

Recent events show that the topic is highly relevant. Statistics shows that individual investors have increased in large amounts in platforms like Robinhood where new accounts have increased by millions within few years (Popper, 2020). Some people call this the "Robinhood effect" and they believe individual investors have a great influence on the stock market in the future. Those who are interested in the stock market are probably aware of the short squeeze of Gamestop stock (GME), which was highly influenced by a Reddit group called r/wallstreetbets. These group of investors relied mostly on word-of-mouth and social media as the primary source of information instead of more traditional sources such as research reports, or conducting their own analysis.

In this thesis I use survey data from a non-representive sample of 140 individuals to investigate factors that correlate with using three different sources of information: (1) word-of-mouth, social media, and other press (Source_F); (2) banks, investing experts/analysts, and specialized press (Source_B); (3) company reports, and annual general meetings (Source_C). Specifically, I explore if the level of overconfidence and individual's age make a difference to the information sources used by the investor. Therefore, the objective of the study is two-fold: (1) identifying whether overconfidence is associated with the sources of information used by investors; (2) identifying whether an investor's age and education influence their choice of used sources of information.

1. LITERATURE REVIEW

In this literature review I go through the main variables of the study and point out what prior studies have pointed out about the topic. I also present theroretical background for consumer's reasons to search for information. Hypotheses will be presented in the end of this section.

1.1. Theoretical background for search for information

The understanding of consumer's reasons to search for information is important will support the study of sources of information used by investors. Srinivasan (1990) studied the decision-making process of consumers from the perspective of external search for information. The three major theoretical perspectives introduced fit perfectly to my study as they explain why investors search for information. The theoretical perspectives are economics, psychological and consumer information-processing perspective.

1.1.1. Economics perspective

The economics perspective is focused on the costs and benefits view of information search in understanding why consumers search for information. Srinivasan (1990) said that when we equate the expected returns of search and expected costs, we find the optimum amount of search. This equation shows why consumers search behavior differentiates. If the investor has higher cost of information search, they will search lesser amount of information. Butters (1978), Rothchild (1978), Salop (1976), and Stiglitz (1979) provide models of economics literature.

Utility maximizer concept is accepted in the field of economics and the logical reasoning to find the equilibrium of cost and returns is justifiable. What one needs to realize is that any criteria may be considered as the objective of utility maximization. The decision maker could value, for example, the lowest price or best quality as the most important factor. Computational advantage may be provided by taking the price in to account alone, but it can be seen as a simplification of

the marketplace. The concept of utility maximization leads to generalization no matter the number of factors you base it on (Srinivasan 1990).

Kohn and Shavell (1974) introduced a term called reservation utility. The limit between "stop searching" and "continue the search" is the utility level which forms the reservation level. Consumers have different reservation utilities for the same products and individuals could have differences in their reservation utility when aspects like time pressure is added. The reservation utility changes when search for information increases and it is not explained why consumers have differences in their reservation utilities. The theory expects that the individual with higher switch point will search more information than the one with lower (Srinivasan 1990). Ratchford (1982) tried the economic cost-benefit framework and showed that the empirical results support the claim where higher reservation utility and lower costs induce higher amounts of search than lower reservation costs and higher search costs.

Parsimony and meaningfulness of the cost-benefit framework are the main advantages of the economic perspective. The outcome of potential benefits and related costs is search activity. The understanding of external search behavior for information can be supported by scrutinizing the determinants of the perception of benefits and costs. By considering the differences and preferences of individuals search activities that impact economic and psychological perceptions of benefits and costs, we can understand why consumers search for information (Srinivasan 1990).

1.1.2. Psychological perspective

Motivation has been the dominating factor in the psychological approach. Howard and Sheth (1969) use attention as the motivational ground of search. Complex stimulus-ambiguity-arousal relationship controls attention. Later Howard added short-term memory and long-term memory to this earlier stimulus-response framework of overt search. The short-term memory processes information from the environment for the long-term memory to use when it is necessary (Srinivasan 1990).

According to Lin (2002) motivation influences the direction and the intensity of the behavior and desires to expend effort on a task. Simon (1967) concluded that motivation is a mechanism which creates and controls an organized goal hierarchy. Motivation is seen as the driver for search activity (Srinivasan 1990).

Two individuals could have different motivation levels to search even if they have the same net utility, in case one of them is an optimizer and the other satisfier. Optimizers see it reasonable to continue information search as long as the net utility is positive. Optimizers are also prone to have a higher net utility. Satisfiers are more likely to choose an alternative but satisfactory choice when there could be a choice with higher utility. They may see it not worthwhile to continue the search even if there is a positive net utility. The different motivation to search should be reflected to the fact whether one wants to optimize or satisfice (Srinivasan 1990)

Another aspect to affect the decision-making process is involvement with the product class (Srinivasan 1990). There have been several studies about the subject (Laurent, Kapferer 1985; Zaichkowsky 1985), but there has not been a jointly agreed opinion about the construct. Srinivasan (1990) noted that the relevance of involvement has been recognized and that with higher involvement comes higher level of motivation to pursue activities.

1.1.3. Consumer information-processing perspective

Main focus is on the memory in consumer information-processing perspective, but it has evolved from the psychological stream. In this approach internal and external are used. A consumer is seen as a goal-oriented and they have so called sub-goals in the various stages of information acquisition. Before a consumer makes a choice, they go through the information acquisition and evaluation process, and memory and external search play a role of sub-components of this stage (Srinivasan 1990).

Theoretically internal search is followed by external search, but the internal search may not be complete before moving into external search. Internal search could be used to acknowledge what needs to be known or internal search could be interrupted because of the lack of information, meaning that internal and external search may be intertwined. This process of going back and forth between internal and external search keeps going until sufficient amount of information has been found and the consumer is ready to make a decision (Srinivasan 1990).

The information-processing theory adds to the basic choice models of consumers that it accounts processing capacity constraints, which takes into account the before assumed view that consumers are capable of handling any amount and quality of information. These processing constraints are inherent in humans, which makes this information-processing perspective so important. Heuristics

helps to avoid extensive information-processing, which leads to lesser processing stress by decreasing the needed amount of information that should be handled (Srinivasan 1990).

Memory influences external search, but the total relationship between internal and external search is not completely understood. Consumers use memory as the inventory of previously acquired information. When consumer is making a new purchase, they recognize the limitations of their processing capacity and the effect of previously acquired information, which could be useful and relevant. Prior knowledge has significant influence on what information the consumer is searching and how they process it. Experience is another dimension of prior knowledge and it has been found to have a negative relationship with external search (Swan 1969; Newman, Staelin 1971). If the consumer has had negative prior experience with a brand, they are more likely pass on that alternative and vice versa with a positive experience. Experience may increase confidence in a consumers' decision-making (Srinivasan 1990).

Schmidt and Spreng (1996) show that the ability to search affects the consumer's information search. They believe that higher level of perceived ability to search will increase external information search activity. Schmidt and Spreng (1996) present three factors which form the perceived ability to search. Those factors are educational level, objective product knowledge, and subjective product knowledge. All these factors have an increasing effect on one's ability to participate in external search (Schmidt, Spreng 1996).

To summarize the theoretical background of consumer's external information search, the main reasons for search are one's costs and benefits of search, their motivation, prior knowledge and experience, and their ability to search for information. The understanding of these factors helps us to understand why investors are searching for information prior their decision to purchase investing instruments. This knowledge gives us the ability to find reasons for the investor's choice of used information sources.

1.2. The dependent variable

The sources of information are related to amount of information gathered and investor's trading frequency (Abreu, Mendes 2012). Abreu and Mendes found connections between investor's

overconfidence level and the sources of information. They show that overconfident traders trade less when they receive information from friends and family.

Information availability may be different since Internet has made it easy to access information in seconds. Mezick (2002) revealed that in the beginning of 21st century investors mainly used newspapers, magazines, and annual reports as information sources for investing purposes. She showed that library use is significant, but Internet has brought the usage down. It is reasonable to think that in the past 20 years Internet has increased it's share as an information provider significantly, but the main effect seems to be that information gathering process is faster. Lewis, Mimura, Mauldin, Rupured and Jordan (2008) said that Internet was not seen as a significant or most reputable source of financial information in their research. The media and internet can be seen as biased information (Lewis *et al*. 2008).

Veronesi (2000) states that it is clear that the prices in the stock market react to new information and tries to find if there is a linkage between information quality and assets returns. Veronesi's model suggests that imprecise information could lead to negative expected excess returns when there is a high risk aversion. Epstein and Schneider (2008) found that investors dislike assets which have poor information quality. Epstein and Schneider provided data which indicates that investor's demand higher excess returns when the stock has low quality of information.

Penning's (2011) view is that investor's use certain sources when they have specific information that needs confirming. In a survey survey made in 1973 used by Penning was shown that more than 60% of investors saw stockbrokers and investment advisors as the most important source of information. Newspapers, magazines and relatives were used as a source but not so often. Company annual reports were seen as the most important less 5% of the time. The accessibility to information via Internet could have changed the significance of different information sources. Many companies have their financial information in their website which makes them easier to access these days, but many investors may see the investment in to this information too costly (it takes a lot of time to analyse company reports carefully) and for that reason rely on experts and analysts.

In a more recent research Piñero-Chousa, Vizcaíno-gonzález and Pérez-Pico (2017) studied the influence of social media to investor's stock market activity. They found that social media influences investor's decision making which results in fluctuations in market risk. In a way social

media works in a similar way with word-of-mouth information. Ivković and Weisbenner (2007) found evidence that an increase of stock purchases from one field made by a neighbour is connected to an increase in a household's own stock purchases in the same field.

1.3. The main independent variables

Tauni, Rao, Fang and Gao (2017) present that investor's personality affects the key information sources used by the investor. They say that financial advisors increase trading frequency within investors who show openness, extraversion, and neuroticism. Durand, Newby and Sanghani (2008) presented that personal traits have an influence on the sources used by the investor. One behavioral factor is overconfidence which has brought forth different opinions on how it affects the investor's performance and actions. Loibl and Hira (2011) noted in their research that risk tolerance and self-confidence could make differences in information search strategies.

One of the main attributes that most have agreed upon is that overconfident investors overstate the quality of their knowledge (Glaser, Weber 2007; Odean 1998; Statman, Throley, Vorkink 2006). Overconfident investors acquire too much information on rely heavily on it, but they respond to the same incentives as rational investors (Guiso, Jappelli 2006). Ko and Huang (2007) found that overconfidence may have a positive impact on market efficiency by increasing information acquisition and that way increasing price quality. Success makes an investor to become overconfident and that when they age, overconfident investors who do not perform well will lose their wealth and confidence (Gervais, Odean 2001).

Higher level of education leads to an increase in the amount of information searched (Claxton, Fry, Portis 1974; Loibl, Hira 2011; Newman, Staelin 1972; Schaninger, Sciglimpaglia 1981). Parents have an important role in the level of their children's financial knowledge and financial management (Lewis *et al*. 2008). Talpsepp, Liivamägi and Vaarmets (2020) say that strong mathematical and academic skills lead to better investment profits and outperformance when other factors like experience and income are considered. They show that investors with higher education have moderate trading frequency and their performance is associated more by finding better trades than any other trading strategy.

Bucher-Koenen and Ziegelmeyer (2013) present that lower financial literacy leads to lower returns in the long run because they are less likely to own risky assets. Van Rooij, Lusardi, and Alessie (2011) show similar results. They see that lack of literacy leads to lower participation in the stock market. Explanation provided by Bucher-Koenen and Ziegelmeyer (2013) is that when a person has good cognitive abilities and knowledge, they face lower cost of information acquirement.

Xia, Wang and Li (2014) show the positive correlation between financial literacy overconfidence and stock market participation. They present in their study that consumers need to have good financial knowledge to participate in the stock market and they must be confident in the level of their knowledge.

Young and old people process information differently and use different sources of information (Phillips, Sternthal 1977). Older people are able to process lesser amounts of information, but they are better at separating relevant information from irrelevant (Schaninger, Sciglimpaglia 1981). Cole and Balasubramanian (1993) found in their study that elderly people search information less intensely and less accurately. They said that in search the ability to process information could be explanatory factor for age differences.

DaSilva and Giannikos (2006) suggested that an investors age is connected to their risk-aversion and that age affects the equity premium and the consumption patterns significantly. Morin and Suarez (1983) concluded that risk-aversion does increase with age, but life cycle seems to have a greater significance on risk-aversion with those investors whose wealth is between \$12,500 and \$100,000.

1.4. Hypotheses

The literature goes through main variables used in my thesis. I have two objectives that I focus my work on, which were mentioned in the end of the introduction. Based on the literature and my objectives, the following hypotheses are presented:

H1 – Overconfidence has a negative association with the likelihood of using word-of-mouth, social media, and other press as information sources

The literature shows that overconfident people trust their own information more and that they overstate the quality of their own information (Glaser, Weber 2007; Odean 1998). Overconfident investors believe that they are performing better than average investors. This indicates that they believe their own information to be better than their friend's or other's. Abreu and Mendes (2012) said that overconfident investors trade less when they receive their information from friends and family. The aim of this study is to check if overconfidence is a significant factor for investor's choice for sources of information when other factors are similar. For example, if there are two investors with similar age, education, and income levels but different confidence levels, whether they have differences in their sources of information.

H2 – Young and less educated are positively associated with the likelihood of using sources such as social media, friends and relatives, and negatively associated with primary sources (i.e., company reports)

High level of education has been connected to better performance and amount of information collected (Talpsepp, Liivamägi, Vaarmets 2020). Lower financial literacy is said to lead to lower return in the long run because of higher probability of owning riskier assets (Bucher-Koenen, Ziegelmeyer 2013). For me this indicates that people with lower education are trusting the information of others and that they are not likely to invest in finding information from more official sources.

Age seems to have an influence on how and how much one can process information (Phillips, Sternthal 1977). Young people are not believed to be as good as older people in identifying quality of information. Age also seems to have effect on risk-aversion (DaSilva, Giannikos 2006). Risk-aversion is said to increase with age. My belief is that younger investors are more likely to endure more risk but that this increased risk is not connected with finding information from primary sources and that younger people believe social media and their friends enough to use them as their main sources of information in investing.

2. RESEARCH METHODOLOGY

2.1 Research approach

A deductive approach was chosen for this research since it is based on previously researched information. Deductive research aims to test if a theory or generalisation applies to a specific occasion (Hyde 2000). Hyde explains that deductive approach is general for quantitative research and that qualitative researchers tend to prefer inductive approach, but both fields do use deductive and inductive approaches. The general agreement is that deductive approach starts with theory, continues with hypothesis development, followed by observation or test, and ends with confirmation or rejection. Soiferman (2010) concluded that the word quantitative could be replaced by the word confirmatory. Inductive approach would not be appropriate for this research because the aim is not to form new theories or hypotheses, but rather confirm those which have been examined before.

In a research it is important to consider the validity and reliability of the methods used. The validity tests if the questionnaire is truly measuring what it is supposed to and if it measures those aspects accurately. The questionnaire in this study was conducted by using previous studies that test the same factors. This gives us an understanding that it does test what it is supposed to. To test that the questionnaire is understood correctly, I asked three volunteers to try the test and explain how they understood the questions. The explanations were homogeneous which gave me the confirmation needed that the questionnaire was understood in the intended way. The questionnaire was acknowledged by the supervisor before publishing.

Fink and Litwin (1995) describe reliability as "a statistical measure of how reproducible the survey instrument's data are". This means that if you would use the same method to the same people, you should produce the same results as before. To improve the reliability, I made the questionnaire in a way that participants could see the whole questionnaire after the demographic questions so they

were able to estimate how long the questionnaire would take. When participants are able to see the end, they may be more willing to finish the questionnaire.

2.2. Data collection and sampling

I collected primary data by a questionnaire using mostly Abreu's and Mendes' (2012) questionnaire questions as a reference and source. Some knowledge base questions were collected from Chen's and Volpe's (1998) analysis of personal financial literacy of college students. The questionnaire was made via Google Forms and distributed via online platforms. I used my personal Facebook, LinkedIn and Shareville accounts where I asked for volunteers to answer my questionnaire. The questionnaire was shared into several personal WhatsApp groups where I asked from friends and family if they wanted to participate. The problem of biased participating was considered, recognizing that close relatives may think participation more as a necessity rather than voluntary. All participants received the same information before answering. Information about the estimated time how long it would take and what the questionnaire was for was given in the description of the posts. I described the questionnaire to be totally confidential and that the subject is sources of information used by consumers. The questionnaire started with demographical questions and asked for investing participation before going for the second part. The second part was different for non-investors and investors, but both tests examined the participant's confidence level and sources of information used for investing purposes. If the participant started the questionnaire and decided not to participate after all, the already answered questions were not reported by the Google Forms system.

The sampling method chosen for this research was non-probability voluntary response sampling. This method is inexpensive and gave the possibility to collect data in a relatively short period of time. By considering the given time frame and budget, the chosen method was seen as the most suitable. My ethical point of view for this method was that by giving the participants the chance to volunteer, they would give true answers. The target was to gain 100 investors to participate and some non-investors to support the data. The data was collected online and shared in my personal social media platforms. The platforms used were Facebook, LinkedIn, WhatsApp and Shareville. The data is cross-sectional data since it is collected in a specific period of time. The precise time being between 26th of March and 7th of April. Shareville, the platforms mentioned before, is a social media platform for the clients of Nordnet, which is a financial service company.

16

2.3 Structure of the questionnaire

The structure of the questionnaire was made to be simple and easy to follow. The possibility to see the total questionnaire after the demographic question was made to decrease rushed and wrongful answers. Most of the questions were taken from previous studies, and some of them were self-conducted to find the information needed for this study.

The questionnaire started with five basic demographic questions and a question if the participant invests in the financial markets. The five demographic questions included age, gender, the highest level of completed education, average monthly net income, and nationality. The last question of this part asked the volunteer's participation in the financial markets. Depending on their answer, the participant moved to the second part. Investors were given a different second part than the non-investors. This was to make sure the participant would not have to face questions they were not able to answer and to decrease flawed and missing data.

The second part for the investors included a total of 15 questions. The questions tested the participants investing activity, investing experience, investing type, stock diversity, risk-aversion, overconfidence, information search activity, used information sources, and financial literacy. Investing activity was tested by simply asking how often the participant buys or sells financial assets. Investing experience was tested by asking how long the participant has been investing in the financial markets. Investing type was tested by asking how long the participant holds their assets. Diversity was examined by giving the participants a possibility to write how many stocks do they own. The participant was asked to scale their risk-aversion from 1 (risk loving) to 10 (risk-averse). Overconfidence questions tested how the participant compares their own performance on different factors, such as the performance of their friends and relatives. The participant was asked to rank themselves from 1 (worse than the compared factor) to 10 (better than the compared one factor). Information search activity question tested how often the participant searched for investing related information. To gather data about the sources of information the participant uses, they were asked to choose one or more from the following choices: Friends & Family (Word-of-mouth), Banks, Investing Experts/Analysts, Social Media, Specialized Press, Other Press, Company reports/Webpage, Annual General Meetings, and other. There were three questions related to financial knowledge. The participant answered two multiple choice questions and one multiple choice grid question where you had to rank five portfolios in terms of expected returns. The last

question for investors asked how confident they were that they answered all the three financial knowledge questions correctly. This was to test their confidence level after a short quiz.

The second part for non-investors had a total of eight questions. The reason for the questions was to provide additional data that could support and bring more information related to the study. The non-investors were asked about if they saved money in any form, their interests in financial market participation, and reasoning to not participate currently. They were also asked to answer questions related to confidence, risk-aversion, and sources of information used if they decided to invest in the financial markets.

2.7 Sample Description

The questionnaire received a total of 140 participants. From all the participants 109 (78%) invested in the financial markets and 31 did not. The 109 responds from investors were used for the logistic regression model. The non-investor information will be used as an additional data to support the study. The descriptive statistics of the total sample can be found in Appendix 1. Detailed information about the variables can be found in Appendix 2.

2.7.1 Socio-economic characteristics

The sample consists of 100 male participants (71%) and 40 female participants (29%). This makes the sample not gender-balanced but should not affect the studied hypotheses significantly. The logistic regression analysis had males referred as 1 and females as 0 (variable *Gender*). The age of the participants varied from 18 to 65. The average age (variable *Age*) was 31,2 years old, and the median was 26. Approximately 35% of the participants were 23 or 24 years old which was expected since the network the questionnaire was presented consisted mainly of people my own age. The level of completed education focused mainly on three groups. From the participants 38 had an upper secondary level education and similarly 38 participants had a master's level education. The largest group was bachelor's level education with 54 respondents. Lower secondary level had 4 participants, all non-investors, and doctorate level participants consisted of 5 respondents. In the logistic regression education was divided in to two groups which were lower (variable *Edu_L*) and higher (variable *Edu_H*) level of education.
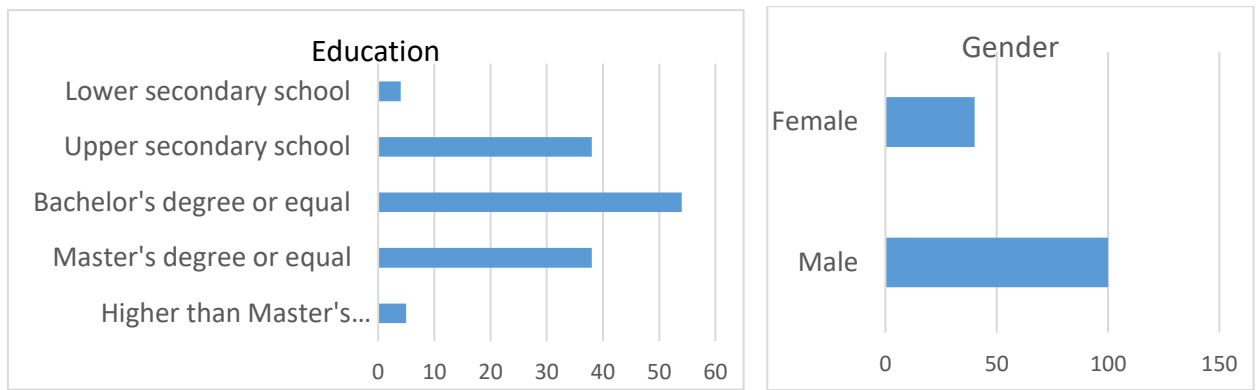
Figure 1. Level of education and gender distribution of the sample
Source: Kallio (2021), author's questionnaire

Participants were asked to choose their average monthly net income level from four different groups. Groups "0-1500" and "1501-2500" were combined into one group in the logistic regression (variable *Inc_L*) and groups "2501-4000" and "4000 or more" into a second group (variable *Inc_H*). *Inc_L* consisted of 52,1% of the participants and *Inc_H* included 47,9% of the participants. From the 140 participants 137 were Finnish, 2 British, and 1 Turkish. Nationality was not seen as a significant factor because most of the participants came from the same country, and nationality was not used in the logistic regression model.

2.7.2 Investors

Investing activity of the participants was tested. 46,8% of the participants said that they invest occasionally. The next option after occasionally was once a month, meaning that close to half of the participants make investments less than once a month. In the logistic regression participants were divided into two groups. Those who invest monthly or more frequently (variable *Activity_H*) and those who trade occasionally (variable *Activity_L*). Investing experience was tested on three levels which were less than two years (variable *Exp_1*), two to five years (variable *Exp_2*), and more than five years (variable *Exp_3*). 48,6% of the participants were include in the highest experience level category (*Exp_3*), 30,3% were in the second group (*Exp_2*), and 21,1% were in the least experienced group (*Exp_1*). The participants were asked how long they hold their assets. Four categories were provided: less than a month, 1-12 months, 1-3 years, and more than 3 years. 62,4% chose more than 3 years, 24,8% hold their assets one to three years, and 12,8% hold their assets for less than year but more than a month. The sample did not have any investors who hold their assets for less than a month. These groups were presented as those who hold assets for less than a year (variable *Inv_L*) and those who hold their assets for more than a year (variable *Inv_H*).

Stock diversity (variable *Diversity*) provided information about if the participant had diversified their portfolio. The mean for *Diversity* was 11,2 and median 9.8 participants did not provide clear information about their diversification. This information was treated as missing values. Risk aversion (variable *Risk_ave*) was tested on a one (risk loving) to ten (risk hating) linear scale. Overconfidence (variable *Overconf*) was tested with four overconfidence questions which gave an average level of overconfidence for the participant. The overconfidence was tested with linear scale from 1 (not confident) to ten (highly confident). The average overconfidence level was 6,61 and the median was 6,75.

Financial literacy was tested with three investing related questions. If the participant answered a question correctly, they were given value 1, and if they answered incorrectly, they were given the value 0. The first question (variable *Fin_Lit_1*) had a correct answer rate of 45%, the second question (variable *Fin_Lit_2*) had a rate of 22,9%, and the third one (variable *Fin_Lit_3*) 41,3%. Confidence level after the questions was tested as well. The second confidence level (variable *Post_Conf*) had a mean of 4,63 and a median of 4. The second confidence level was tested similarly with one (not confident) to ten (highly confident) linear scale. The second confidence level had a clear difference to the first confidence level. The participant was asked to provide information about how often they look for financial information. Four categories were given which were occasionally, monthly, weekly, and daily. These group were combined into two groups for the logistic regression. The first group (variable *Info_Sea_L*) consisted of 36,7% of the participants, and the second group (variable *Info_Sea_H*) consisted of 63,3% of the participants. The sources of information were tested by giving participants choose one or more option from the following option: friends and family (word-of-mouth), banks, Investing experts/analysts, social media, specialized press, other press, company reports/webpage, annual general meetings, and other. A total of eight answers in the other option mainly included sources which were already provided leading to that they did not need to be included in the logistic regression as the participant had chosen category. In the logistic regression the sources were combined as follows:

- (variable *Source_F*) word-of-mouth, social media, and other press
- (variable *Source_B*) banks, investing experts/analysts, and specialized press
- (variable Source_C) company reports, and annual general meetings
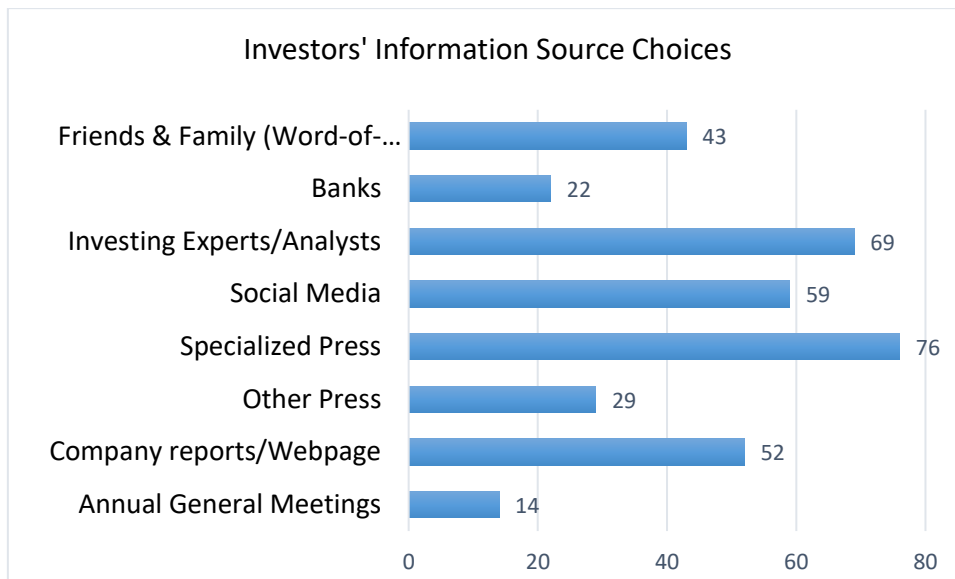
Figure 2. Information source choices of the investors
Source: Kallio (2021), author's questionnaire

2.7.3 Non-investors

Non-investors were asked to provide information about if they had any saving habits. 83,9% of the non-investors informed about saving money in some form. Non-investor's risk aversion was tested similarly as in the investor questionnaire. The participant chose from 1 (risk loving) to ten (risk hating). The mean level of risk aversion was 5,55 which was higher than the investor average. Non-investor overconfidence was tested and showed no clear difference between the first overconfidence level of investor average with a mean of 6,61 on a one (low confidence) to ten (high confidence) linear scale. Non-investors were asked to choose from the same sources of information the ones they would use in case they decided to invest. The non-investors had similar choices with investors, but company reports were chosen only by few of the participants.

Non-investor sources of information choices

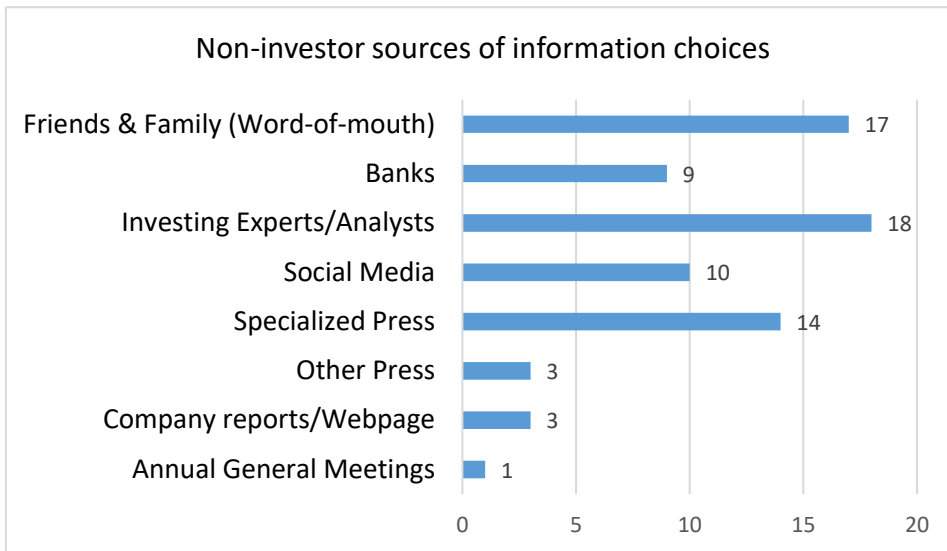| Source	Value
Friends & Family (Word-of-mouth)	17
Banks	9
Investing Experts/Analysts	18
Social Media	10
Specialized Press	14
Other Press	3
Company reports/Webpage	3
Annual General Meetings	1

Figure 3. Information source choices of the non-investor
Source: Kallio (2021), author's questionnaire

3. EMPIRICAL RESULTS

3.1 Results for the regression

I prepared six models for all the dependent values. In the logistic regression the dependent values were the three information source groups (*Source_F, Source_B, and Source_C*). The first models (model 1, model 7, and model 13) simple demographic factors were tested. The first models included *Age* and *Gender* variables. The second models (model 2, model 8, and model 14) included education (*Edu_3*) and investing experience (*Exp_3*). The third models (model 3, model 9, and model 15) added *Diversity* and *Risk_ave* variables, which were predicted to show the influence of risk related variables. The fourth models (model 4, model 10, and model 16) added the influence of investor's trading activity and type. The added variables were *Activity_L*, *Inv_H*, and *Info_Sea_H*. The fifth models (model 5, model 11, and model 17) added overconfidence variables *Overconf* and *Post_Conf*. These variables tested the hypothesis related to investor overconfidence. In the last models (model 6, model 12, and model 18) income and financial literature variables were added to check if these variables add any variation in the results.

The regression shows the dependent value on the top left corner. In the same row it shows the model number. Below the dependent value, the variables are named. Below the variables are shown N which shows the models' number of accounted participants. As eight participants decided to not state or stated incorrectly their *Diversity* value, they were excluded from the third to sixth models. Each model shows two values for each variable included in the model and the significance level. The first value on the left is odds ratio. This value shows the relative odds that the outcome of interest will happen if the exposure to the given variable occurs. Odds ratio higher than one means that exposure to this variable is associated with higher probability of outcome and odds ratio lower than 1 means that exposure to the variable is associated with lower odds of outcome. Standard error (Std. err) shows the average distance of the observed values from the regression line. The smaller the value is, the closer the observation values are to the regression line. The stars next to the standard error values represent significance of the variable.

One star represents significance level of 10%, two stars 5%, and three stars 1%. The adjusted R-squared shows the power of the regression models. If the value increases, it means that the added variables make the model more meaningful and vice versa when it decreases.

The results of the first model of *Source_F* indicate that age is highly significant variable to explain the use of word-of-mouth, social media, and other press as sources of information. The results show that as age increases the probability of using these sources decrease. The significance of age decreases in the next models but stays significant on some level in all the models except in model 2. Gender does not show any significance in any of the models. In this study gender is not seen as a significant factor to explain investors' sources of information choice.

The second model does not give any of the variables clear significance. The model also is the only one that does not give a minimum of 10% significance level for the variable *Age*. The added variables *Edu_H* and *Exp_3* tested if higher level of education or longer investing experience add any explanation to the choice of used information sources. This model differs from the following models by not showing *Exp_3* variable having any significance level. The following four models give investing experience a 5% significance level. The results in the models 3 to 6 show that investors with high level of experience are less likely to use the information sources included in *Source_F* in their information gathering process.

The third model adds the variables *Diversity* and *Risk_ave*. This model has the highest adjusted R-squared value referring it being the most meaningful model. Diversity has a 10% significance level in models 3, 5, and 6. Model 4 does not give any significance level for *Diversity*. The results indicate that having a more diversified portfolio, an investor is more likely to use the information sources in *Source_F*. The significance of *Risk_ave* increases in all the models, gaining the highest level of significance in models 5 and 6. The results suggest that higher level of risk aversion led to lesser likelihood of using *Source_F* information sources.

The fourth model gives results that the frequency of investing and time a investor averagely holds their assets do not have an effect on the investor's use of the referred information sources. High level of information search activity on the other hand does show significance in models 4 and 6. This means that those who actively search for investing information, are more likely to use the information sources of *Source_F*.

Table 1. Logistic regression results of *Source_F*

Source_F	Model 1			Model 2			Model 3			Model 4			Model 5			Model 6		
	Odds	Std. Err		Odds	Std. Err		Odds	Std. Err		Odds	Std. Err		Odds	Std. Err		Odds	Std. Err	
Age	0,946	(0,021)	***	0,966	(0,024)		0,951	(0,029)	*	0,950	(0,031)	*	0,943	(0,034)	*	0,929	(0,039)	*
Gender	1,461	(0,601)		1,794	(0,631)		1,340	(0,725)		1,359	(0,746)		1,631	(0,868)		1,032	(0,988)	
Edu_H				0,457	(0,558)		0,389	(0,629)		0,369	(0,695)		0,313	(0,765)		0,276	(0,891)	
Exp_3				0,413	(0,626)		0,218	(0,770)	**	0,137	(0,870)	**	0,139	(0,994)	**	0,117	(1,075)	**
Diversity							1,074	(0,038)	*	1,056	(0,038)		1,070	(0,041)	*	1,071	(0,041)	*
Risk_ave							0,719	(0,169)	*	0,706	(0,175)	**	0,580	(0,209)	***	0,515	(0,239)	***
Activity_L										1,177	(0,698)		1,017	(0,744)		0,814	(0,767)	
Inv_H										1,599	(1,286)		1,542	(1,386)		1,930	(1,455)	
Info_Sea_H										4,072	(0,798)	*	3,966	(0,913)		5,053	(0,982)	*
Overconf													0,983	(0,418)		1,009	(0,442)	
Post_Conf													0,703	(0,176)	**	0,632	(0,200)	**
Inc_L																1,781	(0,960)	
Fin_Lit_1																3,565	(0,914)	
Fin_Lit_2																1,461	(0,962)	
Fin_Lit_3																0,614	(0,813)	
N	109			109			101			101			101			101		
Adjusted R-squared	2,3 %			3,0 %			10,4 %			7,6 %			9,6 %			4,4 %		

The fifth model adds the overconfidence variables *Overconf* and *Post_Conf*. The general overconfidence questions do not show clear significance for the dependent variable. This could indicate that the average confidence level of investors does not influence sources such as word-of-mouth and social media. When examining the Post_Conf variable we find reverse results. After answering a few finance related questions, investors re-evaluated their confidence. This confidence level shows that those who were confident about answering correctly to all the questions, were less likely to use the information sources included in *Source_F*.

The results of the sixth model indicate that the financial literacy and income level only decrease the meaningfulness of the model. None of the variables *Inc_L, Fin_Lit_1, Fin_Lit_2,* or *Fin_Lit_3* have significance on the studied information sources of *Source_F*. The financial literacy variables are believed to be meaningful for influencing the *Post_Conf* variable rather than the model itself.

Summarizing the dependent variable *Source_F*, it can be said that the set hypotheses are confirmed in some level but not fully. Hypothesis 1 (H1) predicted that higher level of overconfidence decreases the probability of an investor to use word-of-mouth, social media, and other press as sources of information. The results show that the *Post_Conf* variable lowered the probability of using the mentioned sources, but *Overconf* variable did not show clear results for supporting the hypothesis 1. Considering the variables related to hypothesis 2, *Age* did show significance for *Source_F*, but no sufficient findings to support that education influences the information sources in *Source_F*.

Table 2 shows the results of dependent variable *Source_B*. This variable includes banks, investing experts/analysts, and specialized press information sources. This logistic regression was conducted in the same way as the *Source_F* regression. As we can see from the adjusted R-squared values, this regression did not give meaningful results. There are only two independent variables that showed significance. *Risk_ave* showed significance in all the models it was included (model 9 through to model 12). This result suggests that those with higher risk aversion are less likely to use the sources of information in *Source_B*. The significance level of *Risk_ave* was 5% in all the models except model 11 where the significance was on a level of 10%. The other variable showing significance is *Fin_lit_3*. The result indicates that those with some level of financial knowledge are more likely to use *Source_B* information sources. *Fin_Lit_3* was only included in the last model (Model 12) and showed significance level of 5%. In a way this could be expected. Those

Table 2. Logistic regression results of *Source_B*

Source_B	Model 7		Model 8		Model 9			Model 10			Model 11			Model 12		
Variable	*Odds*	*Std. Err*	*Odds*	*Std. Err*	*Odds*	*Std. Err*		*Odds*	*Std. Err*		*Odds*	*Std. Err*		*Odds*	*Std. Err*	
Age	0,973	(0,023)	0,981	(0,027)	0,981	(0,029)		0,985	(0,029)		0,987	(0,031)		0,958	(0,036)	
Gender	2,607	(0,609)	2,641	(0,626)	1,820	(0,671)		1,796	(0,672)		1,716	(0,666)		1,323	(0,719)	
Edu_H			0,457	(0,609)	0,520	(0,643)		0,519	(0,667)		0,510	(0,683)		0,385	(0,799)	
Exp_3			1,095	(0,645)	0,974	(0,690)		0,865	(0,720)		0,795	(0,739)		0,651	(0,889)	
Diversity					1,007	(0,030)		0,998	(0,033)		0,994	(0,034)		0,996	(0,036)	
Risk_ave					0,741	(0,147)	**	0,745	(0,149)	**	0,762	(0,154)	*	0,708	(0,176)	**
Activity_L								0,711	(0,646)		0,754	(0,657)		0,727	(0,734)	
Inv_H								0,735	(1,191)		0,737	(1,203)		0,864	(1,294)	
Info_Sea_H								1,365	(0,665)		1,383	(0,708)		1,908	(0,774)	
Overconf											1,104	(0,338)		1,025	(0,357)	
Post_Conf											1,065	(0,134)		1,093	(0,147)	
Inc_L														2,896	(0,884)	
Fin_lit_1														1,083	(0,726)	
Fin_Lit_2														1,009	(0,820)	
Fin_Lit_3														5,637	(0,747)	**
N	109		109		101			101			101			101		
Adjusted R-squared	-1,5 %		-4,2 %		-4,7 %			-10,9%			-15,1 %			-16,8%		

associated with information sources that often provide knowledge and information directly about financial topics may have more financial knowledge than those who use *Source_F* information source, for example. This is not straight forward tested fact, but rather more of a suggestion for future testing. The summary for *Source_B* is that in this study the model is not seen as meaningful or reliable. There are some indications that level of risk aversion and financial knowledge could affect the investor's used sources of information.

A logistic regression for *Source_C* was made as well. The models had similar results with *Source_B* by having negative adjusted R-squared values, indicating that the models had variables which do not explain the dependent value. The regression found a total of four variables showing significance. The variables were *Exp_3*, *Activity_L*, *Info_Sea_H*, and *Fin_Lit_1*. The significance and result in *Exp_3* suggest that those with higher level of experience in investing are more likely to use *Source_C* information sources than those with less experience. The significance was on 5% level in models 14, 15, and 16 and on 10% level in models 17 and 18. *Activity_L* showed significance of 5% in models 16, 17, and 18. The result shows that investors with lower frequency of trading are more likely to use company reports and annual general meetings as information sources. Those with high level of trading frequency may not have the time to go through lengthy and large amounts of data you gain from the sources in *Source_C*. *Info_Sea_H* showed significance in the same models as *Activity_L*. Investors who gather more information are more likely to use *Source_C* information sources in investing. *Fin_Lit_1* showed significance in the last model (model 18). The result suggest that higher level of financial knowledge would lead to increased level of *Source_C* usage. The logistic regression used for *Source_C* unfortunately does not show meaningfulness and for that reason it is chosen not to be given high value. The significant variables support the theories about consumers information-processing.

3.2 Discussion

In the discussion part I will go through how the hypotheses connected with the results, about the significance of the logistic regressions, and suggest some ideas for future studies.

The hypotheses were conducted in contrast with previous studies and findings. The first hypothesis predicted that overconfidence is associated with lower probability of using word-of-mouth, social media, and other press as sources of information in investing related activities. The *Post_Conf*

Table 3. Logistic regression results of *Source_C*

Source_C	Model 13 Odds	Std. Err	Model 14 Odds	Std. Err	Model 15 Odds	Std. Err		Model 16 Odds	Std. Err		Model 17 Odds	Std. Err		Model 18 Odds	Std. Err	
Age	1,009	(0,017)	0,987	(0,021)	0,987	(0,023)		0,984	(0,024)		0,993	(0,025)		0,983	(0,027)	
Gender	1,367	(0,484)	1,162	(0,499)	0,893	(0,549)		0,892	(0,569)		0,857	(0,587)		0,617	(0,636)	
Edu_H			1,178	(0,451)	1,247	(0,480)		1,288	(0,522)		1,159	(0,530)		0,970	(0,587)	
Exp_3			2,730	(0,445)	2,924	(0,476)	**	3,052	(0,500)	**	2,543	(0,519)	*	2,847	(0,596)	*
Diversity					1,023	(0,023)		1,025	(0,025)		1,023	(0,025)		1,021	(0,026)	
Risk_ave					0,856	(0,105)		0,872	(0,111)		0,892	(0,117)		0,906	(0,129)	
Activity_L								2,660	(0,497)	**	3,112	(0,521)	**	3,227	(0,551)	**
Inv_H								1,356	(0,657)		1,727	(0,695)		1,465	(0,727)	
Info_Sea_H								2,775	(0,509)	**	2,543	(0,526)	*	3,258	(0,566)	**
Overconf											1,407	(0,246)		1,552	(0,268)	
Post_Conf											0,965	(0,095)		0,921	(0,104)	
Inc_L														0,681	(0,590)	
Fin_lit_1														3,504	(0,543)	**
Fin_Lit_2														0,614	(0,579)	
Fin_Lit_3														1,262	(0,504)	
N	109		109		101			101			101			101		
Adjusted R-squared	-3,6 %		-2,4 %		-2,4 %			-1,7 %			-3,2 %			-3,2 %		

variable showed that overconfidence plays a role in the fact if an investor uses these *Source_F* information sources. The results confirm the hypothesis and is connected to the previous studies. The *Overconf* variable on the other hand did not show significance in the models. This could indicate that participants did not show their actual confidence level in the general questions used for overconfidence. The results of *Overconf* were not found being in line with previous studies.

The second hypothesis predicted that those investors with lower level of education and age are more likely to use *Source_F* and less likely to use *Source_C*. The table 4.1.1. shows that age shows significance for the related information sources. The results show that older investors are less likely to use *Source_F* which supports the second hypothesis. Education on the other hand did not show significance in the models and cannot be accepted. Related to the education variable, the specialization was not checked. If education speciality would have been tested the results could have been different. For example, a finance major could have shown difference comparing to a healthcare major. The logistic regression did show significance on the level of investing experience the investor has. This may indicate that education has low or no effect on used information sources, at least when compared to the level of experience one has. The other part of the second hypothesis is rejected because the *Source_C* (Table 4.1.3.) regression showed no support for it. Investors tend to use *Source_C* information sources when they are more actively looking for information and have higher experience level. Testing the education specializations could have fitted better for this as well. Investors with finance specialization are probably introduced to company report analysis which could lead to higher level of using them in personal investing related activities. Age showed no significance on the subject either. With larger sample and higher distribution results could have shown more accurate results.

The logistic regressions had dispersion and two of the dependent variables did not show significance or meaningfulness. This in itself was a good finding too. The questionnaire and variables were conducted by using previous studies and using the variables and factors they found meaningful. This study does show quite clearly that certain factors previously used seem to have no significance on the investor's choice of used information sources. It is unclear if better or significantly different results would be gained when the same models would be tried with a larger and clearly distributed sample. The overall results in this study do show that investors behavioral factors do affect the used information sources. This is connected to the previous studies meaning that this study did in fact follow the previous findings.

For future studies I would suggest to consider examining the information sources in a way that their importance to the investor is accounted as well. In my study only used information sources were asked for, but this did not give data on which the investor uses, trusts, or values the most. For example, using a linear one to ten scale for finding the importance of a certain source of information for an individual investor could be used. An interesting point of view for a similar study would be to study the used information sources and how the investors react to the found information. This could add greater value for the analysists and other investors why certain information causes a reaction and how to predict market reaction.

CONCLUSION

The thesis examined the sources of information used by investors. Most of the investors have preferences which sources they use, and this study focused on finding if overconfidence, age, and education influence the investor's information gathering process. The information sources studied were divided into three groups with the choice made by the author with motivation the sources in the same group being similar in some level.

The study had two objectives which were:
1. to identify if overconfidence creates differences in information sources used by investors
2. to identify if age and level of education affects the sources used by investors

The hypotheses conducted from the viewed literature and given aims were:
1. Overconfidence has a negative association with the likelihood of using word-of-mouth, social media, and other press as information sources
2. Young and less educated are positively associated with the likelihood of using sources such as social media, friends and relatives, and negatively associated with primary sources (i.e., company reports)

The results partly support the previously studied findings and theories made. Investor's overconfidence after answering to three finance related questions was found to have significance on the used information sources. For that reason, the first hypothesis was accepted. Overconfident investors are less likely to use information sources such as word-of-mouth, social media, and other press. The second hypothesis was rejected on the part of young and less educated investors having a lesser chance of using company reports and annual general meetings as sources. The regressions did not show any significance for these factors. Education was suggested to be examined on speciality level in the future studies. On the other hand, young and less educated were found be more prone to use word-of-mouth, social media, and other press as their investing information sources.

This thesis added to the literature by providing more information on the factors affecting investors' choice of used sources. It also provided ideas for future research. This and future studies of the addressed subject could support investors analysis of the movements of the financial markets.

LIST OF REFERENCE

Abreu, M., & Mendes, V. (2012). Information, overconfidence and trading: do the sources of information matter?. *Journal of Economic Psychology*, *33*(4), 868-881.

Bucher-Koenen, T., & Ziegelmeyer, M. (2014). Once burned, twice shy? Financial literacy and wealth losses during the financial crisis. *Review of Finance*, *18*(6), 2215-2246.

Butters, G. R. (1978). Equilibrium distributions of sales and advertising prices. In *Uncertainty in Economics* (pp. 493-513). Academic Press.

Chen, H., & Volpe, R. P. (1998). An analysis of personal financial literacy among college students. *Financial services review*, *7*(2), 107-128.

Claxton, J. D., Fry, J. N., & Portis, B. (1974). A taxonomy of prepurchase information gathering patterns. *Journal of consumer research*, *1*(3), 35-42.

Cole, C. A., & Balasubramanian, S. K. (1993). Age differences in consumers' search for information: Public policy implications. *Journal of consumer research*, *20*(1), 157-169.

Durand, R. B., Newby, R., & Sanghani, J. (2008). An intimate portrait of the individual investor. *The Journal of Behavioral Finance*, *9*(4), 193-208.

Epstein, L. G., & Schneider, M. (2008 cvb). Ambiguity, information quality, and asset pricing. *The Journal of Finance*, *63*(1), 197-228.

Fink, A., & Litwin, M. S. (1995). *How to measure survey reliability and validity* (Vol. 7). Sage.

Gervais, S., & Odean, T. (2001). Learning to be overconfident. *the Review of financial studies*, *14*(1), 1-27.

Glaser, M., & Weber, M. (2007). Overconfidence and trading volume. *The Geneva Risk and Insurance Review*, *32*(1), 1-36.

Guiso, L., & Jappelli, T. (2007). Information acquisition and portfolio performance.

Howard, A. J., & Sheth, J. N. (1969). The theory of buyer behavior. New York: John

Hyde, K. F. (2000). Recognising deductive processes in qualitative research. *Qualitative market research: An international journal*.

Ivković, Z., & Weisbenner, S. (2007). Information diffusion effects in individual investors' common stock purchases: Covet thy neighbors' investment choices. *The Review of Financial Studies*, *20*(4), 1327-1357.

Ko, K. J., & Huang, Z. J. (2007). Arrogance can be a virtue: Overconfidence, information acquisition, and market efficiency. *Journal of Financial Economics*, *84*(2), 529-560.

Kohn, M. G., & Shavell, S. (1974). The theory of search. *Journal of Economic Theory*, *9*(2), 93-123.

Laurent, G., & Kapferer, J. N. (1985). Measuring consumer involvement profiles. *Journal of marketing research*, *22*(1), 41-53.

Lewis, J. K., Mimura, Y., Mauldin, T., Rupured, M., & Jordan, J. (2008). Financial information: is it related to savings and investing knowledge and financial behavior of teenagers?. *Journal of Financial Counseling and Planning*, *19*(2).

Lin, Q. (2002). *Consumers' information search when making investment decisions* (Doctoral dissertation, uga).

Loibl, C., & Hira, T. K. (2011). Know your subject: A gendered perspective on investor information search. *Journal of Behavioral Finance*, *12*(3), 117-130.

Mezick, E. M. (2002). Investing wisely: Individual investors as information seekers. *Journal of Business & Finance Librarianship*, *7*(1), 3-23.

Morin, R. A., & Suarez, A. F. (1983). Risk aversion revisited. *The journal of finance*, *38*(4), 1201-1216.

Newman, J. W., & Staelin, R. (1971). Multivariate analysis of differences in buyer decision time. *Journal of Marketing Research*, *8*(2), 192-198.

Newman, J. W., & Staelin, R. (1972). Prepurchase information seeking for new cars and major household appliances. *Journal of Marketing Research*, *9*(3), 249-257.

Odean, T. (1998). Volume, volatility, price, and profit when all traders are above average. *The journal of finance*, *53*(6), 1887-1934.

Penning, T. (2011). The value of public relations in investor relations: Individual investors' preferred information types, qualities, and sources. *Journalism & Mass Communication Quarterly*, *88*(3), 615-631.

Phillips, L. W., & Sternthal, B. (1977). Age differences in information processing: a perspective on the aged consumer. *Journal of Marketing Research*, *14*(4), 444-457.

Piñeiro-Chousa, J., Vizcaíno-González, M., & Pérez-Pico, A. M. (2017). Influence of social media over the stock market. *Psychology & Marketing*, *34*(1), 101-108.

Popper, N. (8. July 2020). *The New York Times*. Retrieved April 27, 2021, from The new York Times:https://www.nytimes.com/2020/07/08/technology/robinhood-risky-trading.html

Ratchford, B. T. (1982). Cost-benefit models for explaining consumer choice and information seeking behavior. *Management Science*, *28*(2), 197-212.

Rothschild, M. (1978). Models of market organization with imperfect information: A survey. In *Uncertainty in Economics* (pp. 459-491). Academic Press.

Salop, S. (1976). Information and monopolistic competition. *The American Economic Review*, 240-245.

Schaninger, C. M., & Sciglimpaglia, D. (1981). The influence of cognitive personality traits and demographics on consumer information acquisition. *Journal of Consumer research*, *8*(2), 208-216.

Schmidt, J. B., & Spreng, R. A. (1996). A proposed model of external consumer information search. *Journal of the academy of Marketing Science*, *24*(3), 246-256.

Simon, H. A. (1967). Motivational and emotional controls of cognition. *Psychological review*, *74*(1), 29.

Soiferman, L. K. (2010). Compare and Contrast Inductive and Deductive Research Approaches. *Online Submission*.

Srinivasan, N. (1990). Pre-purchase external search for information. *Review of marketing*, *4*, 153-189.

Statman, M., Thorley, S., & Vorkink, K. (2006). Investor overconfidence and trading volume. *The Review of Financial Studies*, *19*(4), 1531-1565.

Stiglitz, J. E. (1979). Equilibrium in product markets with imperfect information. *The American economic review*, *69*(2), 339-345.

Swan, J. E. (1969). Experimental analysis of predecision information seeking. *Journal of Marketing Research*, *6*(2), 192-197.

Talpsepp, T., Liivamägi, K., & Vaarmets, T. (2020). Academic abilities, education and performance in the stock market. *Journal of Banking & Finance*, *117*, 105848.

Tauni, M. Z., Fang, H. X., & Gao, M. (2017). Does investor personality moderate the relationship between information sources and trading behavior? Evidence from Chinese stock market. *Managerial Finance*.

Van Rooij, M., Lusardi, A., & Alessie, R. (2011). Financial literacy and stock market participation. *Journal of Financial Economics*, *101*(2), 449-472.

Veronesi, P. (2000). How does information quality affect stock returns?. *The Journal of Finance*, *55*(2), 807-837. Wiley & Sons, Inc.

Xia, T., Wang, Z., & Li, K. (2014). Financial literacy overconfidence and stock market participation. *Social indicators research*, *119*(3), 1233-1245.

Zaichkowsky, J. L. (1985). Measuring the involvement construct. *Journal of consumer research*, *12*(3), 341-352.

APPENDICES

Appendix 1. Descriptive Statistics

Variable	Mean	Median	S.D.	Min	Max
Age	31,2	26	10,9	18	65
Gender	0,721	1	0,45	0	1
Education					
Lower secondary School	0	0	0	0	0
Upper Secondary School	0,271	0	0,446	0	1
Bachelor's or equal	0,386	0	0,489	0	1
Master's or equal	0,271	0	0,446	0	1
Higher than Master's	0,0357	0	0,186	0	1
Average Monthly Net Income					
0-1500	0,264	0	0,443	0	1
1501-2500	0,257	0	0,439	0	1
2501-4000	0,307	0	0,463	0	1
4001 or more	0,171	0	0,378	0	1
Investing Activity					
Occasionally	0,468	0	0,501	0	1
Once a Month	0,44	0	0,499	0	1
Once a Week	0,0459	0	0,21	0	1
Several Times a Week	0,0459	0	0,21	0	1
Investing Experience					
Less than two years	0,211	0	0,41	0	1
Two to Five years	0,303	0	0,462	0	1
More than Five years	0,486	0	0,502	0	1
Investor type					
Holds assets less than a Month	0	0	0	0	0
Holds assets for 1-12 Months	0,128	0	0,336	0	1
Holds assets 1-3 years	0,248	0	0,434	0	1
Holds assets more than 3 years	0,624	1	0,487	0	1
Stock Diversity	11,2	9	10	0	50
Risk Aversion	4,82	4	2,1	1	10

Variable	Mean	Median	S.D.	Min	Max
Overconfidence	6,61	6,75	1,2	3,5	9,25
Information search activity					
Not at all/ occasionally	0,156	0	0,364	0	1
Monthly	0,211	0	0,41	0	1
Weekly	0,339	0	0,476	0	1
Daily	0,294	0	0,458	0	1
Variable	Mean	Median	S.D.	Min	Max
Banks	0,202	0	0,403	0	1
Investing Experts/Analysts	0,642	1	0,482	0	1
Social Media	0,55	1	0,5	0	1
Specialized Press	0,706	1	0,458	0	1
Other Press	0,266	0	0,444	0	1
Company reports/Webpage	0,477	0	0,502	0	1
Annual General Meetings	0,128	0	0,336	0	1
Financial literacy 1	0,45	0	0,5	0	1
Financial literacy 2	0,229	0	0,422	0	1
Financial literacy 3	0,413	0	0,495	0	1
Post Literacy Confidence	4,63	4	2,73	1	10
Non-Investor saving activity	0,839	1	0,374	0	1
Non-Investor risk aversion	5,55	6	2,03	1	9
Non-Investor Confidence 1	6,61	7	2,03	1	10
Non-Investor Confidence 2	0,645	1	0,486	0	1
Non-Investor Confidence 3	0,258	0	0,445	0	1
Non-Investor Friends & Family (Word-of-mouth)	0,548	1	0,506	0	1
Non-Investor Banks	0,29	0	0,461	0	1
Non-Investor Investing Experts/Analysts	0,581	1	0,502	0	1
Non-Investor Social Media	0,323	0	0,475	0	1
Non-Investor Specialized Press	0,452	0	0,506	0	1
Non-Investor Other Press	0,0968	0	0,301	0	1
Non-Investor Company reports/Webpage	0,0968	0	0,301	0	1
Non-Investor Annual General Meetings	0,0323	0	0,18	0	1

Appendix 2. Description of variables

Variable	Description	Measurement level
Age	Age of the participant	Text slot where participant enters a number
Gender	Gender of participant	Choice between Male, Female, Prefer not to say, and other
Edu_H	Testing the level of completed education. Edu_H included Master's degree and above levels	Choice between lower secondary school. Upper secondary school, Bachelor's or equal, Master's or equal, and higher than Master's degree
Inc_L	Question about the average net monthly income of the participant. Inc_L included choices 0-1500, and 1501-2500.	Choice between 0-1500,1501-2500,2501-4000, and 4001 or more
Activity_L	Testing how often the participant buys or sells financial assets. Those who invest occasionally.	Choice between occasionally, once a month, once a week, and several times a week
Exp_3	Testing how long the participant has been involved with investing. Those who have invested more than 5 years.	Choice between less than two years, two to five years, and more than five years
Inv_H	Testing how long the participant usually holds the owned financial assets. Inv_H includes those who hold their assets more than 3 years,	Choice between less than a month, 1-12 months, 1-3 years, and more than 3 years
Diversity	Question about how many different assets the participant owns	Text slot where participant enters a number
Risk_ave	Testing the level of risk, the participant is ready to take	Linear scale from 1 (risk lover) to 10 (highly risk averse)
Overconf	Testing the participant's level of confidence	Average of four overconfidence questions asked from the participant, all of the questions had a linear scale from 1 (low overconfidence) to 10 (high overconfidence
Overconfidence 1	When you are debating with your friends or relatives about any subject, how often you are the one who is correct?	Linear scale from 1 (I am usually wrong) to 10 (I am the one who is correct)
Overconfidence 2	How do you rate your own knowledge of financial assets and markets on a 1-10 scale (1 = very low, 10 = extremely high)	Linear scale from 1 (very low) to 10 (extremely high)

Overconfidence 3	Compared to others with similar backgrounds as you, how profitable do you see your own assets?	Linear scale from 1 (Not profitable at all) to 10 (Most profitable)
Overconfidence 4	Compared to your friends and acquaintances, how well do you see your financial assets performing?	Linear scale from 1 (Not well at all) to 10 (Extremely well)
Info_Sea_H	Testing how often the participant searches for financial information. Info_Sea_H includes those who search information weekly or daily.	Choice between not at all/occasionally, monthly, weekly, and daily
Information sources	Question about which information sources the participant uses to gain information about financial assets. Friends & Family (Word-of-mouth), Social Media, and other press are included in Source_F. Banks, Investing Experts/Analysts, and specialized press are included in Source_B. The rest are in Source_C.	Choosing one or more from the following options: Friends & Family (Word-of-mouth), Banks, Investing Experts/Analysts, Social Media, Specialized Press, Other Press, Company reports/Webpage, Annual General Meetings, and other
Fin_Lit_1	If you invest €1000 today at 4% for a year, your balance in a year will be - higher if the interest is compounded quarterly rather than weekly - higher if the interest is compounded yearly rather than quarterly - €1000 no matter how the interest is computed - higher if the interest is compounded daily rather than monthly - €1040 no matter how the interest is computed	Given value 1 if answered correctly, otherwise given the value 0
Fin_Lit_2	Rank the following diversified portfolios in terms of expected returns (1 = lowest return, 5 = highest return): - A portfolio of stocks that have only AA-AAA rating - Portfolio that follows a Standard and Poor's composite index - A portfolio of long-term United States Government bonds - A portfolio of small firm stocks - A portfolio of investment grade short term bonds of listed companies	Given value 1 if ranked correctly, otherwise given the value 0

Fin_Lit_3	The dividend payout ratio is: - Used to identify the percentage of earnings per common share allocated to paying cash dividends - Is calculated by dividing the amount of the dividends paid by the earnings per share over a given time frame - An indicator used to determine how well earnings support a dividend payment - All of the above - None of the above	Given value 1 if answered correctly, otherwise given the value 0
Post_Conf	Testing the level of participant's overconfidence after the financial literacy questions: How confident are you that you answered all of the three previous questions correctly?	Linear scale from 1 (Not confident) to 10 (Highly confident)
Non-Investor saving activity	Asking a non-investor participant if they invest in any form	Choice between yes and no
Non-Investor risk aversion	Asking a non-investor participant, the level of their risk-aversion	Linear scale from 1 (risk lover) to 10 (highly risk averse)
Non-Investor Confidence 1	Compared to others with similar background as yours, how well do you believe that you will perform in a random knowledge test?	Linear scale from 1 (Not well at all) to 10 (Extremely well)
Non-Investor Confidence 2	If your workplace or school would have to lay off 50% of the personnel and they would keep those with better than average performance, do you keep your position? (Assume that 50% of the people doing the same job as you are doing)	Choice between "Yes, I would keep my position", "I am not sure", and "No, I would probably lose my position"
Non-Investor Confidence 3	When you are debating with your friends or relatives about any subject, how often you are the one who is correct?	Choice between "Most of the time I am the one who is correct", "About half of the time I am correct", and "I am usually wrong"
Non-investor information sources	Question about information sources: In case you would decide to invest, which of the following sources would you use as your main information sources? (choose one or more)	Choosing one or more from the following options: Friends & Family (Word-of-mouth), Banks, Investing Experts/Analysts, Social Media, Specialized Press, Other Press, Company reports/Webpage, Annual General Meetings, and other

Appendix 3. Correlation Matrix

	Age	Gender	Edu_H	Exp_3	Diversity	Risk_ave	Activity_L	Inv_H	Info_Sea_H	Overconf	Post_Conf	Inc_L	Fin_lit_1	Fin_Lit_2
Age	1,00													
Gender	-0,20	1,00												
Edu_H	0,35	-0,05	1,00											
Exp_3	0,40	0,06	0,27	1,00										
Diversity	0,29	0,06	0,05	0,23	1,00									
Risk_ave	0,10	-0,28	0,06	-0,06	0,07	1,00								
Activity_L	-0,02	-0,10	-0,10	-0,19	-0,24	0,01	1,00							
Inv_H	0,10	-0,07	0,28	0,10	-0,08	0,11	0,03	1,00						
Info_Sea_H	0,09	0,12	0,06	0,22	0,27	-0,20	-0,31	-0,06	1,00					
Overconf	-0,13	0,23	0,06	0,23	0,14	-0,31	-0,25	-0,19	0,29	1,00				
Post_Conf	-0,08	0,26	-0,04	0,08	0,09	-0,28	-0,09	-0,06	0,02	0,42	1,00			
Inc_L	-0,41	0,06	-0,45	-0,39	-0,13	-0,03	-0,02	-0,20	-0,04	0,02	0,09	1,00		
Fin_lit_1	0,04	0,25	0,05	-0,04	0,07	-0,12	-0,04	0,07	-0,07	0,02	0,30	-0,03	1,00	
Fin_Lit_2	0,05	0,06	0,21	0,05	0,04	-0,18	-0,10	0,08	0,05	0,12	0,15	-0,01	0,05	1,00
Fin_Lit_3	-0,18	-0,07	0,06	-0,04	-0,08	-0,14	0,07	0,10	0,03	0,04	0,08	-0,11	-0,05	0,14

Appendix 4. The questionnaire

1. What is your age? *

2. What is your gender *

Female
Male
Prefer not to say
Other: …………

3. What is your highest level of completed education? *

Lower secondary school
Upper secondary school
Bachelor's degree or equal
Master's degree or equal
Higher than Master's degree

4. What is your average monthly net income? *

0-1500
1501-2500
2501-4000
4001 or more

5. Your nationality *

…………

6. Do you invest in the financial markets? *

Yes
No

Section 2 only for those who answered "yes" to the previous question.

7. How often do you sell or buy financial assets? *

Occasionally
Once a month
Once a week
Several times a week

8. How long have you been investing in the financial markets? *

Less than two years
Two to five years
More than five years

9. How long do you usually hold your assets? *

Less than a month
1-12 months
1-3 years
More than 3 years

10. How many different stocks do you own? *

…………

11. How do you consider yourself, on a 1-10 scale, regarding the investment in the stock
    market (1 = love risk, 10 = hate risk) *

12. When you are debating with your friends or relatives about any subject, how often you
    are the one who is correct? * (1 = I am usually wrong, 10 = I am the one who is correct)

13. How often do you search investing information? *

Not at all/occasionally
Monthly
Weekly
Daily

14. What are the main sources you use to find the needed investment information? (choose
    one or more) *

Friends & Family (Word-of-mouth)
Banks
Investing Experts/Analysts
Social Media (Facebook, YouTube, Reddit, Shareville, etc.)
Specialized press (Financial Times, Forbes, Kauppalehti)
Other press
Company reports/webpage
Annual General Meetings

Other: …………

15. How do you rate your own knowledge of financial assets and markets on a 1-10 scale
    (1 = very low, 10 = extremely high) *

16. Compared to others with similar backgrounds as you, how profitable do you see your
    own assets? * (1 = Not profitable at all, 10 = Most profitable)

17. Compared to your friends and acquaintances, how well do you see your financial assets performing? * (1 = Not well at all, 10 = Extremely well)

18. If you invest €1000 today at 4% for a year, your balance in a year will be *

higher if the interest is compounded quarterly rather than weekly
higher if the interest is compounded yearly rather than quarterly
€1000 no matter how the interest is computed
higher if the interest is compounded daily rather than monthly

19. Rank the following diversified portfolios in terms of expected returns (1 = lowest return, 5 = highest return) *

A portfolio of stocks that have only AA-AAA rating
Portfolio that follows a Standard and Poor's composite index
A portfolio of long-term United States Government bonds
A portfolio of small firm stocks
A portfolio of investment grade short term bonds of listed companies

20. The dividend payout ratio is *

Used to identify the percentage of earnings per common share allocated to paying cash dividends
Is calculated by dividing the amount of the dividends paid by the earnings per share over a given time frame
An indicator used to determine how well earnings support a dividend payment
All of the above
None of the above

21. How confident are you that you answered all of the three previous questions correctly? * (1 = Not confident, 10 = Highly confident)

Second section for those who answered "No" to the investing participation question.

22. Do you save money in any form? *

Yes
No

23. Which of the following is the most accurate answer regarding your choice not to invest in the financial markets: *

I do not have the extra money
I do not have the time
I do not know how to

I am not old enough to do it on my own (underage)
I do not see it being important/necessary
It seems to be too risky and I do not want to lose my wealth

24. Disregarding the previous answer and assuming knowledge, about how to invest, limits your participation in the financial markets, would you invest in the stock market if someone would teach you how to do it? *

Yes
No
I know how to and I have other reasons not to participate

25. Do you see yourself as a risk-taking person? Answer on a 1-10 scale, 1 = risk lover, 10 = extremely risk-averse *

26. Compared to others with similar background as yours, how well do you believe that you will perform in a random knowledge test? * (1 = Not well at all, 10 = Extremely well)

27. In case you would decide to invest, which of the following sources would you use as your main information sources? (choose one or more) *

Friends & Family (Word-of-mouth)
Banks
Investing Experts/Analysts
Social Media (Facebook, YouTube, Reddit, Shareville, etc.)
Specialized press (Financial Times, Forbes, Kauppalehti)
Other press
Company reports/webpage
Annual General Meetings

Other: …………

28. If your workplace or school would have to lay off 50% of the personnel and they would keep those with better than average performance, do you keep your position? (Assume that 50% of the people doing the same job as you are doing) *

Yes, I would keep my position
I am not sure
No, I would probably lose my position

29. When you are debating with your friends or relatives about any subject, how often you are the one who is correct? *

Most of the time I am the one who is correct

About half of the time I am correct
I am usually wrong

Appendix 5. Non-exclusive licence

A non-exclusive licence for reproduction and publication of a graduation thesis[11]

I Anton Mikael Kallio (*author's name*)

1. Grant Tallinn University of Technology free licence (non-exclusive licence) for my thesis
Investment related information gathering by different age and education groups: The
effect of overconfidence to information sources,

(*title of the graduation thesis*)

supervised by_____ Pavlo Illiashenko _____,
(*supervisor's name*)

1.1   to be reproduced for the purposes of preservation and electronic publication of the
graduation thesis, incl. to be entered in the digital collection of the library of Tallinn University of
Technology until expiry of the term of copyright;

1.2   to be published via the web of Tallinn University of Technology, incl. to be entered in the
digital collection of the library of Tallinn University of Technology until expiry of the term of
copyright.

2. I am aware that the author also retains the rights specified in clause 1 of the non-exclusive
licence.

3. I confirm that granting the non-exclusive licence does not infringe other persons' intellectual
property rights, the rights arising from the Personal Data Protection Act or rights arising from other
legislation.

_____

_____ (date)

---

[1] *The non-exclusive licence is not valid during the validity of access restriction indicated in the student's application
for restriction on access to the graduation thesis that has been signed by the school's dean, except in case of the
university's right to reproduce the thesis for preservation purposes only. If a graduation thesis is based on the joint
creative activity of two or more persons and the co-author(s) has/have not granted, by the set deadline, the student
defending his/her graduation thesis consent to reproduce and publish the graduation thesis in compliance with clauses
1.1 and 1.2 of the non-exclusive licence, the non-exclusive license shall not be valid for the period.*