TALLINN UNIVERSITY OF TECHNOLOGY
School of Business and Governance
Department of Economics and Finance

Leo Estola

Alphabetic Bias in Nordic Stocks

Bachelor's thesis

Programme International Business Administration, specialisation Finance

Supervisor: Pavlo Illiashenko

Tallinn 2022

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 8003 words from the introduction to the end of conclusion.

Leo Estola ……………………………
(signature, date)
Student code: 195106TVTB
Student e-mail address: lestol@ttu.ee

Supervisor: Pavlo Illiashenko, MBA:
The paper conforms to requirements in force

……………………………………………
(signature, date)

Chairman of the Defence Committee:
Permitted to the defence
…………………………………
(name, signature, date)

TABLE OF CONTENTS

ABSTRACT

Alphabetical ordering in the firms tickers has shown a bias towards the firms in the beginning. The first alphabets in firm’s tickers has shown tendency to have more volume than the ones placed in the bottom. Biases can skew markets and it is important to recognize them. Research of alphabetical bias from the United States found out that there was indeed a bias in the U.S. stock market. The findings and the data was from U.S. and it did not include any other stock markets.

The aim of this thesis is to find out if this bias exists in Nordic stock markets as well and what could possibly be the best first letter to name a company. The data is gathered from three years between 2019 and 2022. The coronavirus pandemic happens to be on the same timeline during this study.

This research carries out quantitative methods to evaluate the alphabetic bias. Many regressions are made with different sample sizes ranging from over six hundred to more than three hundred depending on data that was possible to find and by eliminating outliers from 95th and 5th percentile of the data. Several variables are included and analyed to get more accurate outcomes.

The results unfortunately did not show signs of the bias in the beginning of the alphabet. However, the letter M that is in the middle of the alphabet seemed to perform better than the rest of the letters and was only letter that had significancy.

Keywords: Alphabetic bias, Volume, Regression, Nordics

INTRODUCTION

Here, in the introduction, I am going to go through the alphabetic bias in general and pull-out research from relevant studies, tell the methods I used in the process and give an overall structure from this thesis. I am also going to address my research problem and question along with hypothesis.

“Words are the source of all power. And names are more than just collection of letters.”

- Rick Riordan

This quote by Riordan has even deeper meaning than one could assume by the first glance. Humans can see only to a certain extent with a first glance. We’re simple creatures, we can stay focused only for a short period of time. When looking at the grocery store list, most of us remember clearly the first things written in that paper and not the ones in the middle, for example. This same thing was studied by Van Praag and Van Praag (2008) with name ordering on multi-authored academic papers. It concluded that there is an advantage on the early placement of the alphabet. Professor A, who has been a first author more often than Professor Z, will have published more articles and experienced a faster productivity rate over the course of her career because of reputation and visibility. The list goes on. The studies I have gathered in the background section were conducted all around the world, which could mean that this bias is not country specific by any means, it could be global.

Alphabetic bias order in stocks has been researched by the thorough study done by Jacobs and Hillert (2016) in novel of the Review of Finance. The study concluded that the alphabetical ordering tends to provide an advantage to those positioned in the beginning of an alphabetical listing. It is said that the article was first of its kind by the extensivity and implication of this bias in financial markets. Apparently, stocks that appear near the top of an alphabetical listing have about 5-15% higher trading activity and liquidity than stocks that appear in the bottom. Sample was US stocks from the biggest stock exchanges such as NASDAQ and NYSE with over 6000 share available and data from companies all the way from 1983 to 2011. Could this bias be in other stock exchanges as well?

Because it was mentioned in the through study by Jacobs and Hillert (2016) that the trading activity is a huge part of the reason stocks in the US have this alphabetic bias. Therefore, this thesis addresses stock's alphabetic position and its effect on volume (trading activity) in the Nordic stock exchanges from 08/02/2019 to 08/02/2022. My hypothesis is that there is a similar bias as in US stocks and stocks with their tickers placed high in the alphabet have higher volume, since biases are heuristics that affect all the people's subconsciously. Not only this paper finds out whether the alphabetic bias exists, it also tells us what the best and worst first letter is to name a company. My research question for this study is:

1. Is there a benefit on firms' ticker placed in the start of the alphabet?

Biases are crucial to find and study, because it affects the whole economy and market efficiency eventually. The timeline is chosen to examine the effects of coronavirus pandemic as well. To be more precise, I have gathered data from one year before the coronavirus crashed markets in 2020, data one year during the crash and one year after the crash, to compare, I also have returns from 2 years before the pandemic. I will draw conclusions from each year and from all the years together.

Jacobs and Hillert (2016) study suggested that firm's name in an alphabetically ordered list may be used as an instrument for trading activity in future research. Since, that study's main factor affecting the bias was the trading activity, this paper will calculate regression with variables with the first letter of the ticker and the volume and the first letter of the ticker with normalized volume data and market cap data to provide more accurate regression as well. I will be providing a few interesting figures to visualize the most traded firms' first letters and the lowest. The Nordic stock exchanges has much less companies listed in it compared to the US. The original sample consisted of 637 large- mid- and small cap firms from Nordic stock exchanges but it was cut down to 357 in the last few regressions to eliminate the outliers. The data was gathered from Google Finance. The data came with daily volume, but some smaller firms were missing values, the data is calculated as weekly averages to achieve more precise data set.

The next chapters of this thesis are background and conceptual framework where I will lay foundation for my paper, second chapter is data and methods where I tell how I gathered my data for this paper and how I came up with the results for my analysis, next up is the results and discussion where I present the analyzed data and lastly conclusion where I bind together my findings, talk about the limitations affecting the study results and pave a way for further research.

1. BACKGROUND AND CONSEPTUAL FRAMEWORK

Alphabetic bias is a form of behavioral bias. These biases affect people's behavior unconsciously. In this first chapter I will talk about other financial biases that affect our decision making and talk more about alphabetic bias from previous studies conducted to lay foundation for my reason to study this matter.

1.1. Behavioural biases

Study from Said, Asl, Rostami, Gholipour and Gholipour (2011) has done great research on perceptual biases effects on financial decisions. They have written the abstract in a smart and compact way where they say that their study is to "recognize the popular perceptual errors among investors and its connection with their personality". They took a sample size of 200 random investors from Tehran's stock market and collected data was gathered from questions using parametric analysis and correlation to check their hypotheses that were the following: 1st there is a significant correlation between being extroverted and investors' perceptual errors, 2nd there is a significant correlation between conscientiousness and the investors' perceptual errors in stock market, 3rd there is a significant correlation between agreeableness and the investors' perceptual errors in stock market, 4th there is a significant correlation between the neuroticism and the investor's perceptual errors and lastly 5th there is a significant correlation between openness and investor's perceptual errors. This study concluded that there was a strong relation between the investors' personality and the perceptual errors in Tehran's stock market. Four hypotheses from this study were confirmed and one was rejected. Second hypothesis was rejected because there was not apparently correlation between the investors' agreeableness and perceptual errors. By examining these results, it is confirmed that there is relation between extroversion and hindsight, there is reverse relation between dutifulness and randomness, there is a straight relation between neuroticism and randomness and there is a direct relation between openness and hindsight over confidence and that a reverse relation between openness and availability. This tells us clearly that there are biases affecting our behavior.

Next up I am going to go through 17 other financial biases besides alphabetic bias that affects our decision making. The biases are taken from Zahera and Bansal (2017).

Overconfidence. This bias affects to stock market performance forecasts. We, the people, uniformly overestimate our knowledge and our ability to predict on a huge scale. Overconfidence is the difference what people think they know and what they really know, as Dobelli (2014) formulates it.

Disposition effect. This bias was identified by Shefrin and Statman (1985). Investors are rushing to sell their stocks with yield very early to capitalize the gains and tend to hold losing stocks in a fear of losses. People rather avoid losses more than they are willing to realize them. Average investor’s decisions are more based on the fact that they have made profit and not on the losses.

Herding effect. As Shiller (2000) and Kahneman and Tversky (1979) defined it as investors are more willing to follow other’s decisions rather than find information themselves. This causes shocks in the market as deviations in prices from the fundamentals values and could cause reduction in returns.

Mental accounting. This was a work of Thaler (1985) and it concluded that people divide their investments in different portfolios on the grounds of how many mental categories they have. This could lead to non-profitable portfolios only based on the emotions.

Confirmation bias. This was given by Dickens (1978), and it states that people have prejudiced opinions on what they think and rely on this information alone. They then adapt this thinking and try to fit it in the future information as well. This could lead to avoiding information and cause distorted reality of what they know.

Hindsight bias. It was invented by Fischhoff and Beyth (1975) and it means that investors rely on that the phenomenon can be predicted accurately. This causes irrational decisions as one tries to draw up cause and effect relationships that have no factual base.

House money effect. When gamblers are winning hugely, they start to take more risk and become risk-averse Thaler and Johnson (1990)

Endowment effect. Investors are missing profitable investments because they like to hold what they currently have and do not want to switch assets. So, some asset prices stay abnormally low without a real reason Kahneman, Knetsch and Thaler (1990)

Loss aversion. This bias is quite fascinating, and it was given by Benartzi and Thaler (1995). When faced with confirmed profits people tend not to take any risk and but if they are faced with confirmed losses suddenly, they are willing to take risk.

Framing. Same information but framed differently can have a huge impact on how people perceive things. If information is given in a positive manner, investors are not willing to take risk but if it is modified in a negative manner, they are prepared to take risk in a fear of losses Tversky and Kahneman (1981).

Home bias. Investors want to hold domestic firms' stocks even though they are not doing so great rather than buy foreign stocks French and Porteba (1991) and Tesar and Werner (1995).

Self-attribution bias. Founded by Bem (1967, 1972). People think they are so smart, and the success is because of their own traits and when things go wrong, they blame others or factors outside their grasp.

Conservatism bias. Investors are stubborn and only trust their own gathered information and beliefs and are not open for information which might be useful for them Edwards (1982).

Regret aversion. People are scared to fail. Once they fail, they do not want to have the same feeling of failure. One failure could affect their future decisions for no good reason Fishburn (2013).

Recency bias. Freshly news of some events has effect on investors' decision making and they refuse to use information that could be useful for them from the past Heery and Noon (2008)

Anchoring. Investors make their decisions based on the information they have from the past and then base their newly given information around the past information. The information "anchors" around the past information Tversky and Kahneman (1981).

Representativeness. If some event that has occurred in the past has similar features in the present, people may think the same event could happen in the future. History is not a guarantee of the future Kahneman and Tversky (1979).

Overconfidence and return were studied by Dorfleitner and Scheckenbach (2022) in two major used trading platforms in Germany. As the studies from above suggests, people are subject to behavioral biases. This research pointed out that in the trading platforms, there were various irrational factors that significantly linked the overconfidence bias. The trader’s popularity (either measured by the number of followers inside the platform or the net change in invested capital) reveals to be significant factor of this so-called irrational behavior on both trading platforms. Therefore, the authors’ draw up a conclusion that overconfidence of the traders increase when they get more attention as they gain capital inflows. There was negative relationship between trading activity and performance on social trading platforms as Dorfleitner and Scheckenbach (2022) findings suggests.

Kourtidis, Chatzoglou and Sevic (2016) wanted to find out does the role of personality traits in investors trading behavior in Greece. The personality traits used here are overconfidence, risk tolerance and sociability. The study used an innovative integrated model using difficult structural equation modelling analysis to be able to examine them simultaneously as they would occur in the real world. When the results came, the authors found out that the most powerful relationships were between the overconfidence and stock trading performance, frequency, and volume. The results also indicated that overconfidence positively influenced stock trading volume.

As we can see, humans are no way near rational decision makers or act rationally even if they think they do so to call. These behavioral studies acknowledge the issue of non-rationality. We can pull together an argument that the markets we are investing are driven by irrational humans with their own behavioral biases and every person has their own different behavioral biases. This behavioral approach gives this study a great foundation to study more of the biases that are surrounding our everyday life. These biases I found and talked about are just the tip of the iceberg, the most notable and recognized ones. There must be many more factors and biases affecting human behavior we just have not found them all to this day. As the research go deeper and further, we are able to find more irrationality and understand human behavior more thoroughly. The next bias that I am going to go through is the alphabetic bias, which is usually known as more of a psychological bias and it is not researched as much in the financial field.

1.2. Alphabetic bias

Research done by Doelmann, Itzkowitz, Itskowitz and Sardali (2017) studied alphabetic bias in fund allocation decisions plans. They found out that the funds which are listed at the beginning of the plan's menus receive much greater allocations compared to the funds that are listed more at the bottom. This study says that it is not surprising that this phenomenon exists, because the more choices we are faced with, the more our natural instincts and heuristics come to use when it is time to make a decision. Notable to mention is that according to this research, this bias tends to grow as the funds in the plan offering increase. This concluded that investors prone to alphabetic bias, the same factors that apparently biased their allocation choices could lead them to pick funds that improve overall profitability outcomes.

A study was conducted where Jurajda and Munich (2010) found out that when applying to a university and one is being close to the admission margin amongst other applicants, those placed higher in the alphabet enjoy higher chances of admission. It could be explained with Ang, Chua and Jiang (2010) study where preference for A is ingrained in people's minds to evoke a more positive affect than B. The fact that A is preferred over B is typical affect heuristics. Thus, branding consumer products with labels A facilitates marketing and increases selling prices and market valuation.

It really does make a difference on where one is placed in an alphabet. For example, study done by Feenberg, Ganguli, Gaule and Gruber (2015) examined this phenomenon by reviewing consumer responses from National Bureau of Economic Research (NBER). At the start of each week, NBER issues a "New This Week" email, which is a collection of all the new working papers from the past week. Apparently, the email goes to more than 23 000 subscribers inside and outside of the academia and the placed order is based on random factors. Despite the randomized list placement, papers that are listed first week are 30% more likely to be viewed, downloaded, and cited over the next years. Vice versa, lower ranking on the list on the list led to fewer views and downloads, but not cites. Nonetheless, there was recency bias as well, because the freshly listed paper was getting more hits, downloads and cites. This indicates that a "randomized" list can be manipulated by this bias.

Others have also been interested in this subject of the first character determining unfair biases around the globe. Richardson (2008) wanted to examine whether American Journal of

Roentgenology (AJR) reviewers are biased towards first letter of an alphabet with last names. He collected data from database of Editorial Manager. It is a Web-based software used by AJR to review and manage journal production. The names were listed alphabetically. Richardson analyzed this extracted data with R software. During the 224-day sample period, 1 195 manuscripts were submitted to AJR, and 5 825 invitations were sent to a pool of 1 573. Not so shockingly, the trend was downward from A to Z. There was a linear association between the number of invitations and the alphabetic position of the first letter of the reviewers' last names (r=-0.75). The results were statistically significant as well and proved with chi-square goodness-of-fit test. The reviewers with their names starting in the beginning of the alphabet had almost twice as many invitations to review. Richardson argued that this bias is most likely due to the "satisfaction of search" by the assigning editors who invite the reviewers who meet their criteria first.

Einav and Yariv (2006) referenced this bias as "alphabetical discrimination", where they studied a faculty in top 35 U.S. economics departments. Faculty with earlier placement surname initials are significantly more likely to receive tenure at the top ten economics departments, are significantly more likely to become part of the respected association of Econometric Society and even more likely to receive the Clark Medal and the Nobel Prize. Einav and Yariv (2006) also proved that the statistically significant differences remained after the same control for country origin, ethnicity, religion, or fixed departmental fixed assets.

Same kind of confirmation about alphabetical discrimination was conducted by Huang (2014), where author used vast sample size from Web of Science to see if ABC's gets more citations than XYZ's. The research finds the relationship between surname initials and paper citations. Further in the study Huang extends the reference lists and finds out that alphabetic bias is much stronger when reference lists are longer and that the bias is also stronger for papers published freshly. It is a standard to list authors alphabetically to any research papers, so the impact must be huge.

Another study done from alphabetic bias only supports this bias. Stevens and Duque (2019) used 150 000 articles to test whether alphabetizing in-text citations biases readers into eventually citing more articles with authors that name's start in the beginning of the alphabet. Results indicated that surnames in the start of an alphabet were cited way more than those later in it when the journal ordered the same citations alphabetically compared to more objective, chronological or numerical order. This tells an important message: first is not always the best just because one's cognitive process gives more attention to items placed earlier in the list rather than in the later on as Simon (1990) says: “Human short-term memory can hold only a half dozen chunks”.

A psychological study conducted by Carney and Banaji (2012) wanted to see if first is the best. We evidently experience information sequentially according to these researchers. Why is that? Well, they say that there is no rational reason for it. They conducted three social experiments. In the first one they tested the “first is best” hypothesis using three different pairs of stimuli (two male salespersons, two female salespersons and two teams) and for the participants of this experiment were given each item of the pair in a sequence and then they were asked their preference. The procedure in a nutshell went like this: it was a three different choice-pairs and they needed to pick a preference. First were teams they wished to join (“Hadleys” and “Rodsons”), secondly who would they buy car from (“Jim” and “Jon”), thirdly who would they buy a car from (“Lisa” and “Lori”). The sample size for this first experiment was one hundred twenty-three guineapigs. The results were collected, and ANOVA measures taken. The one presented first compared to the second option was significantly chosen more often. Second experiment consisted of choosing a consumer good from two options. Two pieces of similarly packaged and flavored bubble gum. The trick here was that for some groups the authors gave them a time constraint to choose a forced and spontaneous decision and for some groups they gave them time to choose their option. Here were two-hundred seven participants. The results were aligned with the first experiment. Participants from the rapid task chose the first option more often but the participants from the slow choosing process had more even result (51% and 49% in this case). In the third experiment authors wanted to see if showing pictures of two criminals taken from Florida Department of Corrections website, which one should be more worthy of parole and less worthy for parole. The males in the photos were both 29 years old and had committed same violent crimes. They had same outfits and neutral faces. These researchers hired two coders to and modified the pictures from facial expressions extremely negative to extremely positive. Same with attractiveness. In this experiment there were thirty-one participants. Even though the stimuli were negative (prisoners and who should remain in prison), participants automatically associated the first criminal to being more worthy for parole. I guess first is the best after all.

This bias affects also in the national level, which is not fair at all. In the UK 2010 local government elections in the Greater London area Wood, Badawood, Dykes and Slingsby (2011) studies the relationship between candidates’ position on a ballot paper and vote rank was explored with a sample size of 5000 candidates. The results were unbelievable. They show that position bias was significant enough to affect the rank order of candidates and that some candidates that were (currently at the time of the authors wrote this study in 2011) were representing London may have benefitted from this effect. This effect was powerful enough to confer first positioned candidates a six-times advantage over third positioned in the same party. Authors said that there is some evidence as well that the impact of this is sufficient to overcome the voter preference for party.

How about choosing securities? Is there an affect with the letter effect? Knewtson and Sias (2010) wanted to analyze whether asset selection is influenced by the starting letter. It came clear that there is effect for investors to select a security with classical economic theory. As said in the paper: "The number of institutions holding a company's shares is greater for companies whose names begin with more common name letters. In addition, undergraduate students are more likely to select securities for evaluation when the first letter in the company's name matches the first letter of their first name or either their first or last name". Result indicates that there is emotional bias as well as rational stimuli on investors' decisions. From this research, it is said that stocks that begin with the common name letter M, has greater number of institutional shareholders than counterparts that start with the bottom of the alphabet.

Now there is evidence of alphabetic bias in the overall human nature according to these few very throrough researches. The bias did not affect everyone in the studies above, but on average, it had a significant effect that leaned towards recognizing alphabetic bias. But not everyone are affected by the same biases the same way and trying to act rationally can have an opposite effect. It is important to recognize all sorts of biases so we can learn from them and find solutions. If we let all kinds of biases control us, it can cause unfair outcomes as the Wood, Badawood, Dykes and Slingsby (2011) found that by simply ordering ballot papers alphabetically can lead to someone not so competent people to run a town, city or even country. There must be all sorts of cases where this bias has skewed even statistcics to wrong direction.

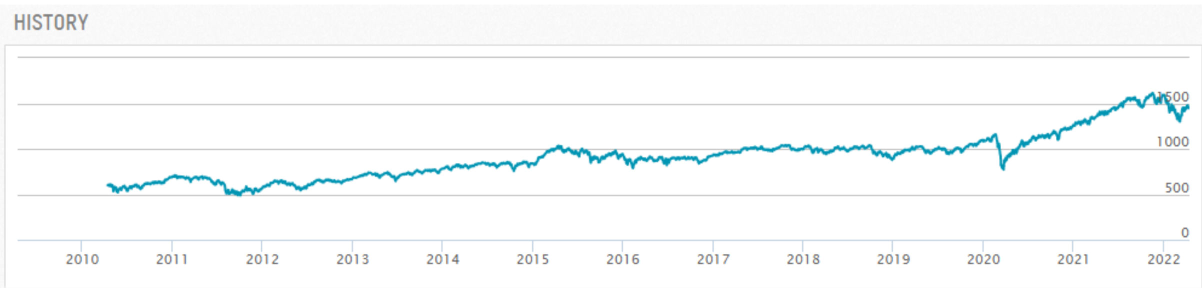
2. DATA AND METHODS

In this chapter I am going to enlighten about my sample size and from where is it collected. I will also tell about the methods on why I chose them and how I used them and the overall process of how the process went from collecting the data to getting results from it.

2.1. Sample data

The data consists of 637 Nordic large-, mid- and small cap firms from the beginning of February 2019 to the beginning of February in 2022. The time period in this study is crucial, since it was chosen from one year before the coronavirus pandemic started to two years after its outbreak. The data is Nordic stock's volume data, and it is gathered from three years. As we can see in the Figure 1, which represents NOMXN120 index development, the stocks started to crash on the 20th of February in 2020. My sample firms' tickers are listed in the appendix in a Google Drive format, where there is all my raw Excel and Google Sheet data.

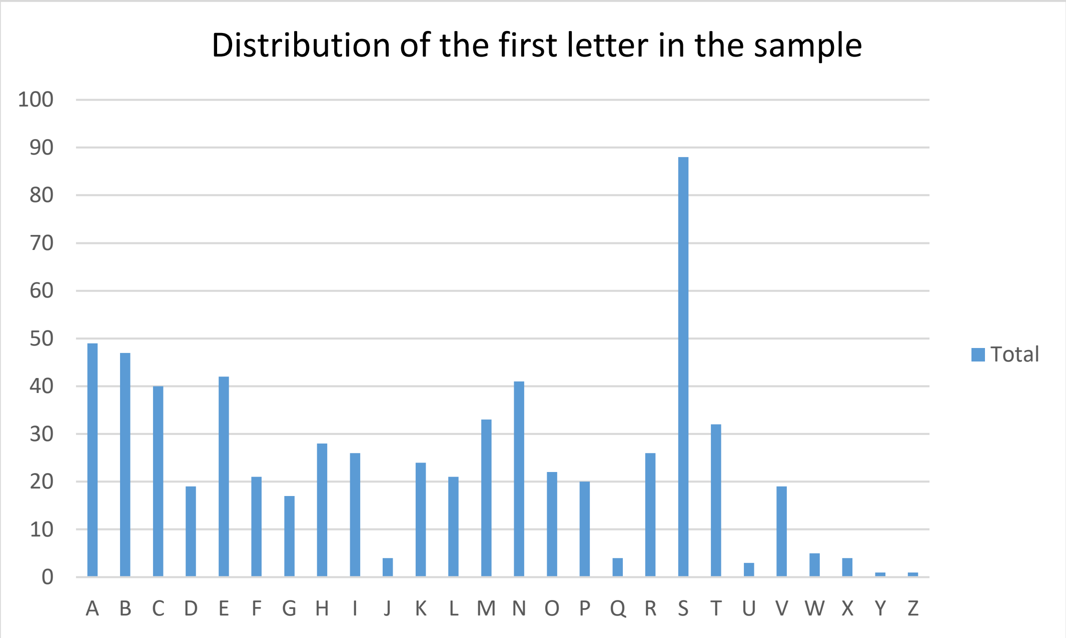
Figure 1: OMXN 120 development from 2010-2022



Source: Nasdaq Nordic (2022)

The data was collected from Google Sheets, because it was the most efficient way to collect this massive data of almost 670 000 cells of volume data. Syntax GOOGLEFINANCE(ticker, [attribute], [start_date], [end_date|num_days], [interval]) was used. Where attribute was volume, start date was 02/08/2019, end date was 02/08/2022 and the interval was weekly. Figure 2 shows each letters' distribution through the alphabet.

Figure 2: Distribution of the first letters



Source: Author’s calculations

2.2. Methods

After gathering all the weekly volume data from Google Sheets, the data was moved to Excel for smoother user experience and calculations. The next step was to calculate the weekly volume sums for the first year, second year, third year and all these year’s sum with the sum function. The data needed to be aligned properly and edited in a way that it was possible to copy and paste the sum function so that the dates would not be messed up. They needed to have same number of rows with the same weekly dates. If there was lacking year or two years of data, it needed to be done by hand. After counting the sums, it was time for sorting the volume sums with the first letter of the ticker. The SORT function was used. Syntax SORT(array;(sort_index);(sort_order);(by_col)). This was sorted for all three years and the sum of all years obviously and the sorting function also organized the data from high to low with the first letters that moved with the volume cell. After that, it was time to analyze the data. Eight pareto charts were made (two for each year). Top 50 biggest volumes and their first letter and top 50 lowest volumes and their first letter.

After getting these pareto charts where it is easily visualized, which letters are doing the best and the worst during all the periods it was time to calculate regression with two variables.

The independent variable was firm tickers' first letter converted into numbers so that A=1, B=2, C=3 and so on in order to perform regression analysis with Excel's data analysis tool, since it would not accept letters as values. Find and replace function was used so that. The dependent variable was the volume sums. I had to transpose the previously listed tickers to columns instead of rows because the regression analysis would not accept 637 columns as the array.

To also provide more accurate data and change these values to common scale, I wanted to normalize this volume data with firms' market cap. Volume data is so much larger than alphabetic numbers, so it makes sense to make the ranges a bit closer. I downloaded the market cap of each firm with the same function in Google Sheets. Now the attribute was market cap. The idea behind this was that I could calculate the average volume for each firm and divide it with the market cap with the corresponding firms. So, with the firms coming from four different countries (Finland, Sweden, Denmark, and Iceland) we obviously had four different currencies. To get the currencies of every ticker my attribute in the function was currency. Now it was easy to convert all the market caps to same currency. I used Euros. I searched up the exchange rates of three currencies on how much is it in Euros and multiplied the market cap with the exchange rate.

I was not satisfied with the results that came with volume and tickers to numbers regression. I needed to have more calculations on this bias. After getting the ratio of average volume and current market cap, I saw the outliers were huge and they had to be removed. 95th and 5th percentile of the ratios were removed to provide more accurate understanding. Now I had new data with no outliers. The number of observations dropped from 637 to 569. I also removed some other data if I did not have enough information about them. In the new sheet I had numbered tickers' vowels as 1 and consonants as 2. There is also return of the stock from 3 years and from 5 years to see if there is any difference and 2 years and return from the 2 years before that was calculated with these two returns. This sheet also includes market returns for each according stock from 2 years before and 3 years during. So, four returns for four country indices (ICEX, OMX HEL 25, OMX STO 30 and OMX COP 25). With these returns I am able to calculate excess return for 2 years before and 3 years during. Returns are a good proxy because stocks that go up and down have more volume. Basically, if stock goes up 15% and market goes up 10% the excess return is 5%. I tried to get beta values as well from the Google Sheets, but for some reason it only gave me about 50 so it made no sense to use betas as proxy. I switched beta values to p/e values, since I had them for 357 companies and high p/e ratio usually means that the company is riskier. There are also columns from first five letters of the tickers as ones and other letters are converted into zeros. I did the same thing with first 10 letters correspondingly.

To do the regression analysis on our proxies and control variables, I needed to download Gretl, it is an open-source statistical package for econometrics. In Excel one can not treat letter as a continuous variable. I converted this Excel file into CSV file. In the Gretl it needs to be clarified that our letters are not numbers. We need to make them dummy variables. After dummifying those letters, we have, in the regression we need to remove the dummy for the letter A, so it serves as the basis for comparison. I did few models with the variables I have. The average volume and market cap ratio was the dependent variable. The proxies were letter position, first 5, first 10 and vowels. Control variables included excess return from 2 years before and during the 3 years, and price to earnings ratio. I put together 9 different models.

3. DISCUSSION AND RESULTS

This part of the thesis goes through firstly the descriptive statistics on the data and regression analysis from the results. I will tell about the important numbers and values that need to be taken in to consideration. This part has many tables and figures that not all could be shown here and they will be found in the appendix.

3.1. Descriptive statistics

According to the pareto charts, 1st years' best letters to name a firm was S, E and I and the worst letters were T, S and F. 2nd year best letters were S, I and N and the worst were S, T and R. 3rd year best letters were N, S and I and the worst were T, B and N. All the years' best letters were S, N and I and the worst were T, F and H as we can see from the figures 6, 7, 8, 9, 10, 11, 12 and 13 in the appendix. The pareto charts shows that the best possible letter to name a firm (highest and lowest volume) is S and the worst is T.

With the correlation matrix we can clearly see that with the volume ratio to market cap the p/e ratio is the most affecting factor with the value of 0,24. Other variables (position, vowel, first 5, first 10, excess return 2 years before and excess return from 3 years during only has values ranging from -0,04 to 0,05. This is a good thing so we can see that price to earnings ratio has correlation with volume. Figure 3 shows the correlations between different variables.

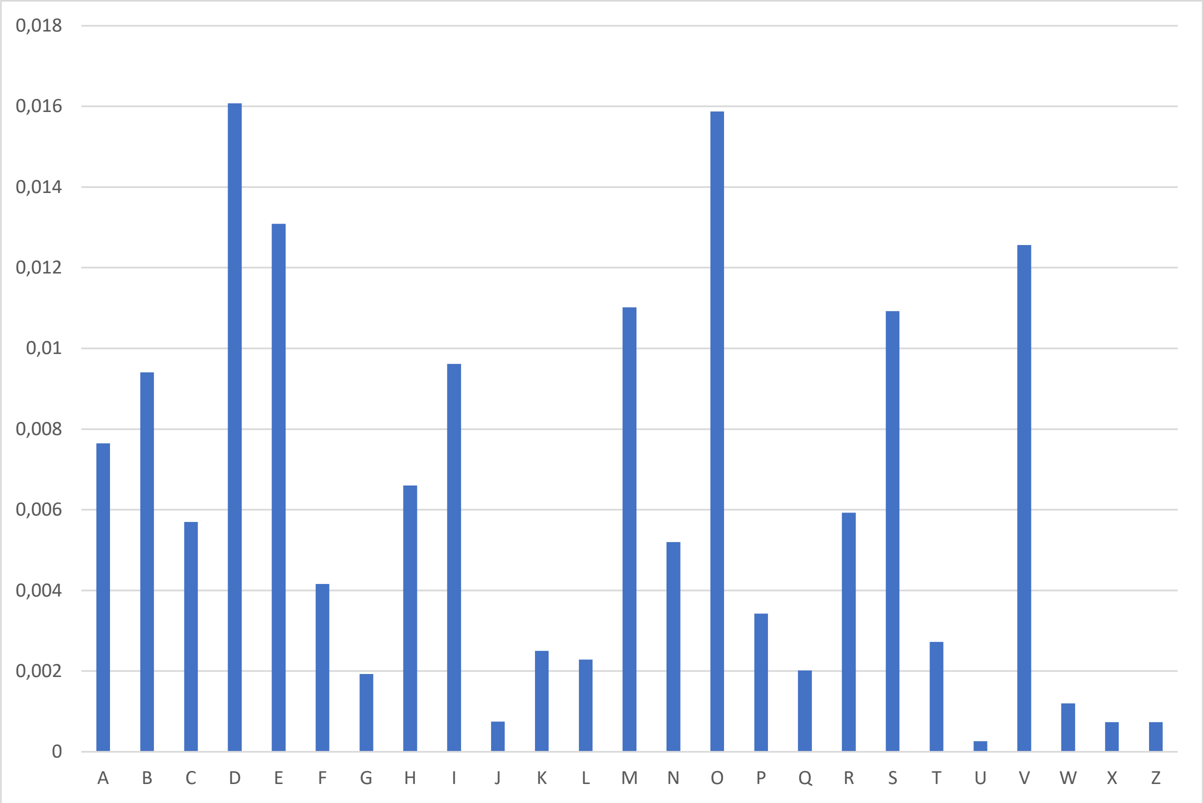
Figure 3: Correlation matrix

	ratio	position	vowel	first.5	first.10	ex.ret.2y.before	ex.ret.3y.during	pe
ratio	1,00							
position	-0,04	1,00						
vowel	0,05	-0,40	1,00					
first.5	0,05	-0,75	0,45	1,00				
first.10	0,00	-0,87	0,42	0,67	1,00			
ex.ret.2y.before	0,01	-0,07	0,01	0,06	0,11	1,00		
ex.ret.3y.during	0,00	0,05	-0,06	-0,04	-0,04	-0,02	1,00	
pe	0,24	0,03	0,11	-0,02	-0,04	0,01	-0,03	1,00

Source: Author's own calculations

This figure has letters used and their average ratio. When grouping the average ratios from first 5 letters, it is clearly lower than the rest of the letters. This is the same with the first 10 letters of the alphabet. Figure 4 visualizes the average ratio with each letter and how it is distributed.

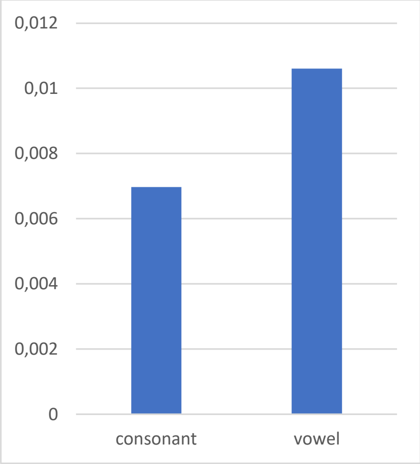
Figure 4: Letters and their average ratio



Source: Author’s own calculations

But surprisingly enough the vowel letters have higher average ratio than consonants. Almost two times more. Figure 5 illustrates the difference between consonants and vowels.

Figure 5: Consonants and vowels average ratio



Source: Author's own calculations

3.2. Regression analysis

The summary from volume and first letter of the ticker regression analysis concluded that in the 1st year from the top 50 the r-squared was 0.57%, 2nd year top 50 r-square was 0.0045%, 3rd year top 50 r-squared was 0.0029%, and all the years top 50 r-squared was 0.0042%. 1st year lowest 50 r-squared was 1.6%, 2nd year lowest 50 r-squared was 3.72%, 3rd year lowest 50 r-squared was 0.22%, and the all years' lowest 50 r-squared was 1.96% as we can see from the tables 1, 2, 3 and 4, 5, 6, 7 and 8 in the appendix. Next up I will go through a more precise analysis from these tables.

In the Regression statistic table 1 first year top 50, the Multiple R equals 0.076 which means there is basically no linear relationship whatsoever. R-squared or the coefficient of determination of 0.57% tells us that the response variable cannot be explained by the predictor variable at all. We do not need the adjusted r-squared since we have only one predictor here. F statistic is 0.27, it is not significant. Significance of F or P-value is 0.6 so this is not statistically significant because it is higher than 0.05.

In the Regression statistic table 2 second year top 50, the Multiple R equals 0.0067 which means there is basically no linear relationship whatsoever. R-squared or the coefficient of determination of 0.0046% tells us that the response variable cannot be explained by the predictor variable at all.

We do not need the adjusted r-squared since we have only one predictor here. Significance of F or P-value is 0.96 so this is not statistically significant because it is higher than 0.05.

In the Regression statistic table 3 third year top 50, the Multiple R equals 0.0054 which means there is basically no linear relationship whatsoever. R-squared or the coefficient of determination of 0,0029% tells us that the response variable cannot be explained by the predictor variable at all. We do not need the adjusted r-squared since we have only one predictor here. Significance of F or P-value is 0.97 so this is not statistically significant because it is higher than 0.05.

In the Regression statistic table 4 all year's top 50, the Multiple R equals 0,0065 which means there is basically no linear relationship whatsoever. R-squared or the coefficient of determination of 0,0042% tells us that the response variable cannot be explained by the predictor variable at all. We do not need the adjusted r-squared since we have only one predictor here. Significance of F or P-value is 0.96 so this is not statistically significant because it is higher than 0.05.

In the Regression statistic table 5 first year lowest 50, the Multiple R equals 0,13 which means there is basically no linear relationship whatsoever. R-squared or the coefficient of determination of 1,6% tells us that the response variable cannot be explained by the predictor variable at all. We do not need the adjusted r-squared since we have only one predictor here. F statistic is 0,78, it is not significant. Significance of F or P-value is 0,38 so this is not statistically significant because it is higher than 0.05.

In the Regression statistic table 6 second year lowest 50, the Multiple R equals 0,19 which means there is basically no linear relationship whatsoever. R-squared or the coefficient of determination of 3,7% tells us that the response variable cannot be explained by the predictor variable at all. We do not need the adjusted r-squared since we have only one predictor here. F statistic is 1,85, it is not significant. Significance of F or P-value is 0,18 so this is not statistically significant because it is higher than 0.05.

In the Regression statistic table 7 for the third year lowest 50, the Multiple R equals 0,047 which means there is basically no linear relationship whatsoever. R-squared or the coefficient of determination of 0,22% tells us that the response variable cannot be explained by the predictor variable at all. We do not need the adjusted r-squared since we have only one predictor here. F statistic is 0,11, it is not significant. Significance of F or P-value is 0,74 so this is not statistically significant because it is higher than 0.05.

In the Regression statistic table 8 for all the year's top 50, the Multiple R equals 0,14 which means there is basically no linear relationship whatsoever. R-squared or the coefficient of determination of 1,2% tells us that the response variable cannot be explained by the predictor variable at all. We do not need the adjusted r-squared since we have only one predictor here. F statistic is 0,96, it is not significant. Significance of F or P-value is 0,33 so this is not statistically significant because it is higher than 0.05.

With the volume regression analysis, it seems that there's no difference in the first letter in Nordic stocks at the top 50, because the regression trendline is pretty much horizontal line, but there is with the bottom 50, since the trendline has upwards trend and therefore bottom letters in the alphabet is highlighted. This could mean that here in the Nordics we have more tickers starting with the bottom of the alphabet.

Regressions results do not tell us much about reality but what we can find with the data we got. One star on the significance means that the p-value is 5-10%, two stars mean that the p-value is 1-5% and three stars means that the p-value is under 1%. The table 9 done with Gretl shows that model 1 with excess returns 2 years before has no significance with our ratio and adjusted r-square equals 0,26%. The number of observations is 569 because of eliminating the 5th and 95th percentile of data. The

Second model from the table 9 has regression from ratio and excess returns from 3 years during. There is still no significance and the adjusted r-square equals 0,019%, even less than before. Observations are 569.

Third model from the table 9 calculates regression from the excess return 2 years before and 3 years during with dependent variable being the ratio obviously. The adjusted r-square is pretty much the same with the value of 0,02% with 569 observations.

With the table 9's last model, model 4, we have ratio as the dependent variable and excess return from 2 years before, excess return from 3 years during and p/e ratio as control variables. Now the observations have dropped to 357 because of the limited p/e data but the adjusted r-square has now gone up to 4,73%. P/e ratio is significant, and the three stars tells us that the p-value is less than 1%. There is clearly correlation.

Table 9: Regression models done with Gretl

	Model 1			Model 2			Model 3			Model 4			Model 5		
	Coef	St. Er	Significance	Coef	St. Er	Significance	Coef	St. Er	Significance	Coef	St. Er	Significance	Coef	St. Er	Significance
const	0,008	(0,001)	***	0,0079	(0,001)	***	0,008	(0,001)	***	0,00451	(0,001)	***	0,004553	(0,001)	***
exret2ybefore	-0,001	(0,001)					-0,001	(0,001)		0,00013	(0,001)				
exret3yduring				-0,001	(0,001)		-0,001	(0,001)		0,00009	(0,001)				
pe										0,00001	(0,000)	***	0,000008	(0,000)	***
Adj. R2	0,26 %			0,019 %			0,02 %			4,73 %			5,26 %		
N	569			569			569			357			357		

Source: Author's own calculations

Moving on to table 10 with 3 models in it. First off, we have ratio as dependent variable and p/e and first 5 as control variables. There's significance with p/e as we found out but not with the first 5. However, the adjusted r-square has gone up to 5,30% with the observations being 357.

Model 2 from table 10 has ratio as dependent variable and p/e and first 10 as controls. No significance found with first 10 and the adjusted r-square has gone down to 5,01% with the same number of observations.

Last model from this table 10, model 3 has ratio as dependent variable and p/e and vowel as control variables. There's no significance with vowel. The adjusted r-square has gone slightly up to 5,05% with 357 observations.

Table 10: Regression models done with Gretl

	Model 1			Model 2			Model 3		
	Coef	St. Er	Significance	Coef	St. Er	Significance	Coef	St. Er	Significance
const	0,0041929	(0,0007225)	***	0,0044287	(0,0008128)	***	0,0044163	(0,0007071)	***
pe	0,0000084	(0,0000018)	***	0,0000083	(0,0000018)	***	0,0000082	(0,0000018)	***
first 5	0,0016575	(0,0015444)							
first 10				0,0003254	(0,0013126)				
vowel							0,0007471	(0,0016379)	
Adj. R2	5,30 %			5,01 %			5,05 %		
N	357			357			357		

Source: Author's own calculations

Regression table 11 with only 1 model holds inside the ratio as dependent variable and p/e plus all the dummified letters excluding the letter A dummy as control variables. There is significance with dummy letter 5, which in our case equals to letter M. Letter M is in the middle of alphabet and therefore we can reject the hypothesis that there would be benefit on placed on the start of the alphabet with significancy of one star equaling 5 to 10% of p-value. If the ticker starts with the letter M, then the dependent variable is 0,0070 higher than if the tickers start with letter A.

Table 11: Regression model done with

	Model 1		
	Coef	St. Er	Significance
const	0,0047	(0,0027)	*
pe	0,0000	(0,0000)	***
Dletter_1	0,0014	(0,0031)	
Dletter_2	0,0002	(0,0038)	
Dletter_3	-0,0047	(0,0043)	
Dletter_4	0,0044	(0,0053)	
Dletter_5	0,0070	(0,0038)	*
Dletter_6	0,0056	(0,0040)	
Dletter_7	0,0007	(0,0046)	
Dletter_8	-0,0038	(0,0038)	
Dletter_9	-0,0024	(0,0042)	
Dletter_11	0,0015	(0,0040)	
Dletter_12	0,0021	(0,0042)	
Dletter_13	-0,0016	(0,0035)	
Dletter_14	-0,0004	(0,0043)	
Dletter_15	-0,0030	(0,0037)	
Dletter_16	-0,0023	(0,0040)	
Dletter_17	-0,0001	(0,0042)	
Dletter_18	-0,0027	(0,0040)	
Dletter_19	0,0025	(0,0123)	
Dletter_20	-0,0028	(0,0048)	
Dletter_21	-0,0036	(0,0060)	
Dletter_22	-0,0045	(0,0089)	
Dletter_23	-0,0088	(0,0089)	
Dletter_25	-0,0044	(0,0089)	
Adj. R2	5,26 %		
N	357		

Source: Author’s own calculations

CONCLUSION

The literature review gave me motivation to chase the alphabetic bias on Nordic stocks, since the studies showed that the bias is proven many times scientifically. Unfortunately, from my data I could not find any signs supporting this bias in stock tickers from 637 stocks firstly 569 and 357 afterwards. I was hoping to see more stronger correlation (much more than under 5,30% r-squared and adjusted r-squared that I got) from the regression analysis. The case could be that many of the firms in the Nordic has more smaller firms compared to U.S. stocks. The stocks in the U.S. are much larger in scale of Nordics. Their share price and spot in the huge indices surely affects the results and they are not recognized universally the same way. They have better liquidity overall, since the markets are larger and there are many huge enterprises recognized all around the globe. My sample data had firms from small-, mid- and large caps in Nordic standards, which are not even close to American sizes.

After doing eight different regression analyses with different time periods in regards of the top 50 and lowest 50 volumes and the sum of all the periods and a few regression analyses with the normalized values as the volume and market cap ratios, I must reject the hypothesis that there would be an alphabetic bias on the first letter of a firm. There was no significant difference according to regression analyses that a firm name from the beginning of the alphabet would have more trading activity. I am using r-squared and adjusted r-squared as the comparison. There is no significance with the r-squared values that would suggest the bias to exist. The highest value for r-squared and in the top 50 was 0.57% and the lowest was 0.0029%. The highest r-squared value in the lowest top 50 was 3.71% and the lowest was 0.22%. Highest value of adjusted r-squared from all the models was 5,30%. This analysis only suggests that it does not matter on how one names the company. However, in the lowest volume regression analysis the results suggests that there is an upward trend the bottom of the alphabet is highlighted. It could mean that here in the Nordics, we have more firms starting with the bottom part of the alphabet. It could make sense in a way that our alphabet has 31 letters in it and the English alphabet has 26. This papers' data had 26 letters in it, because one cannot use letters Š, Ž, Å, Ä & Ö in the stock exchange tickers. Even though this data uses same letters that English alphabet has, it could be integrated in Nordic people’s minds that our alphabet is not ending in letter Z and there’s more to it. We can easily name a company starting in the lowest middle part of our own known alphabet.

Pareto charts visualized the alphabetic volume data wonderfully. The best letter to name a firm according to the volume data descriptive statistics was S and the worst was T. Volume sum data is on the left vertical axis of the chart, the cumulative percentage is on the right vertical axis and the letters are descending order by bars on the horizontal axis. With the regression analysis with the letters being dummy variables the matter was that the letter M would be the best possible letter to name a firm. It was the only letter to have any kind of significancy between all the other letters.

There are some limitations to this study. Firstly, since the daily data tends to have missing values especially for the smaller firms, the data is treated as weekly averages to assure a more balanced data set. Secondly, some of the firms in my sample has not been around for three years. Few of the firms has listed in the stock exchange after February 2019, so there’s only data from one or two years. Thirdly, few firms have also went bankrupt or their shares has been delisted during the period, so there might not be data from the last year or the year before. The names starting with letters Š, Ž, Å, Ä & Ö are treated with S, Z, A, A & O respectively. I had to remove 83 firms from my original 720 sample size, since there were missing values due to reasons mentioned above. It was impossible to calculate precise data with the 83 firms since they did not have enough data for one year. Also, the different regressions used different number of observations if the different variables did not have enough data to be calculated with other data. The outliers of the data were also eliminated in the last part of the regression to ensure more accurate and balanced data.

Ideas for further research. The already existing literature and results of this study could inspire someone to try this same study but a more thoroughly or try the next ideas mentioned below.
I would suggest that for the future research one should check the sectors’ correlation and best & worst letter divided by sectors. See if the countries have different correlation and best & worst letters and see if there is correlation between the three different firm sizes (large cap, medium cap and small cap) and see the best and worst letter from those.

The firms from Nordics are much smaller in size than in the U.S. firms inside the top indices, so they do not have the same liquidity than in the U.S. Many of my sample stocks could be treated as “penny stocks” and therefore there is not much interest in them from the public. The analysis was not as in depth as the big research team’s one, so maybe there could be a bias if done a thorough study

LIST OF REFERENCES

Ang, J., Chua, A. and Jiang, D., (2018). Is A Better than B? How Affect Influences the Marketing and Pricing of Financial Securities. [online] Taylor & Francis Online. Available at: <https://www.tandfonline.com/doi/abs/10.2469/faj.v66.n6.5> [Accessed 19 March 2022].

Bem, D.J. (1967), "Self-perception: an alternative interpretation of cognitive dissonance phenomena", Psychological Review, Vol. 74 No. 3, pp. 183-200.

Bem, D.J. (1972), "Self-perception theory", Advances in Experimental Social Psychology, Vol. 6, pp. 1-62.

Benartzi, S. and Thaler, R.H. (1995), "Myopic loss aversion and the equity premium puzzle", The Quarterly Journal of Economics, Vol. 110 No. 1, pp. 73-92.

Carney D. R. Banaji M. R. (2012) First is best, PLoS One7 , 1 – 5 .

Dickens, C. (1978), "Introduction", in Pattern, R.L. (Ed.), The Pickwick Papers, Penguin Books, ISBN 978-0-415-22233-4.

Dobelli, R. (2014). The Art of Thinking Clearly. USA: Harper Paperbacks.

Doelmann, Itzkowitz, Itzkowitz, Sardali (2017). Alphabeticity Bias in 401(k) Investing.

Dorfleitner, G. and Scheckenbach, I. (2022), "Trading activity on social trading platforms – a behavioral approach*", Journal of Risk Finance, Vol. 23 No. 1, pp. 32-54. https://doi.org/10.1108/JRF-11-2020-0230

Edmund Heery and Mike Noon, (2008) Oxford University Press. A dictionary of human resource management (2ed.).

Edwards, W. (1982), “Conservatism in Human Information Processing (excerpted)”, in Kahneman, D., Slovic, P. and Tversky, A. (Eds), Judgment under Uncertainty: Heuristics and Biases, Cambridge University Press, New York, NY.

Einav, L. and Yariv, L., (2005). *What's in a Surname? The Effects of Surname Initials on Academic Success*. [online] American Economic Assosiation. Available at: <https://www.aeaweb.org/articles?id=10.1257/089533006776526085> [Accessed 19 March 2022].

Feenberg, D., Ganguli, I., Gaulé, P. and Gruber, J., (2017). *It’s Good to Be First: Order Bias in Reading and Citing NBER Working Papers*. [online] MIT Press Direct. Available at: <https://direct.mit.edu/rest/article/99/1/32/58370/It-s-Good-to-Be-First-Order-Bias-in-Reading-and> [Accessed 19 March 2022].

Figure 1: Nasdaq OMX Nordic index (photograph). Retrieved from: http://www.nasdaqomxnordic.com/indexes/historical_prices?Instrument=SE0003270875

Fischhoff, B. and Beyth, R. (1975), “I knew it would happen: remembered probabilities of once-future things”, Organizational Behavior and Human Performance, Vol. 13 No. 1, pp. 1-16

Fishburn, P.C. (2013), The Foundations of Expected Utility, Springer Science & Business Media, Vol. 31.

French, K. and Poterba, J. (1991), “Investor diversification and international equity markets”, American Economic Review, Vol. 81 No. 2, pp. 222-226.

Gretl open-source econometric regressions. Retrieved from: http://gretl.sourceforge.net/?msclkid=390844abd13411ecb17b9a21b82c22fb

Huang W. (2015) ‘Do ABCs Get More Citations than XYZs?’, Economic Inquiry, 53/1: 773–89.

Jacobs, H. and Hillert, A., (2016). Alphabetic Bias, Investor Recognition, and Trading Behavior *. [online] Oxford Academic. Available at: <https://academic.oup.com/rof/article/20/2/693/2461444?login=false#123793163> [Accessed 19 March 2022].

Jeffrey R. Stevens, Juan F. Duque. (2019). Order matters: Alphabetizing in-text citations biases citation rates. Psychonomic Bulletin & Review 26:3, 1020-1026.

Jurajda and Munich (2010). Alphabetical Order effects in School Admissions.

Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision Under Risk. Econometrica, 47(2), 263–291. http://dx.doi.org/10.2307/1914185

Kahneman, D., Knetsch, J.L. and Thaler, R.H. (1990), "Experimental tests of the endowment effect and the coase theorem", Journal of Political Economy, Vol. 98 No. 6, pp. 1325-1348

Knewtson, H. and Sias, R. W., (2010). Why Susie owns Starbucks: The name letter effect in security selection. Journal of business Research, 63(12), 1324-1327

Kourtidis, D., Chatzoglou, P. and Sevic, Z. (2017), "The role of personality traits in investors trading behaviour: empirical evidence from Greek", International Journal of Social Economics, Vol. 44 No. 11, pp. 1402-1420. https://doi.org/10.1108/IJSE-07-2014-0151

Rasoul Said, Hassan Ghalibaf Asl, Mohammad Reza Rostami, Aryan Gholipour, Fattaneh Gholipour (October 2011). International Journal of Economics and Finance Vol. 3, No. 5; Behavioral Finance: The Explanation of Investors' Personality and Perceptual Biases Effects on Financial Decisions.

Richardson, M., (2008). Alphabetic Bias in the Selection of Reviewers for the American Journal of Roentgenology : American Journal of Roentgenology: Vol. 191, No. 6 (AJR). [online] American Journal of Roentgenology. Available at: <https://www.ajronline.org/doi/full/10.2214/AJR.07.3737> [Accessed 19 March 2022].

Shefrin, H. and Statman, M. (1985), “The disposition to sell winners too early and ride losers too long: theory and evidence”, The Journal of Finance, Vol. 40 No. 3, pp. 777-790.

Shiller, R.J. (2000), Irrational Exuberance, Princeton University Press, pp. 149-153

Simon, H. A. (1990). Invariants of human behavior. Annual Review of Psychology, 41, 1–20.

Syed Aliya Zahera and Rohit Bansal, Accepted 6 November (2017). Do investors exhibit behavioral biases in investment decision making? A systematic review.

Tesar, L. and Werner, I. (1995), “Home bias and high turnover”, Journal of International Money and Finance, Vol. 14 No. 4, pp. 467-492.

Thaler, R. (1985), “Mental accounting and consumer choice”, Marketing Science, Vol. 4 No. 3, pp. 199-214

Thaler, R.H. and Johnson, E.J. (1990), “Gambling with the house money and trying to break even: the effects of prior outcomes on risky choice”, Management Science, Vol. 36 No. 6, pp. 643-660.

Tversky, A. and Kahneman, D. (1981), “The framing of decisions and the psychology of choice”, Science New Series, Vol. 211 No. 4481, pp. 453-458.

Van Praag, C. M., and van Praag B. M., (2008). The benefits of being economics professor A (rather than Z). Economica, 75, 782 796.

Wood, J., Badawood, D., Dykes, J. & Slingsby, A., (2011). BallotMaps: Detecting name bias in alphabetically ordered ballot papers. *IEEE Transactions on Visualization and Computer Graphics,* 17(12), 2384-2391.

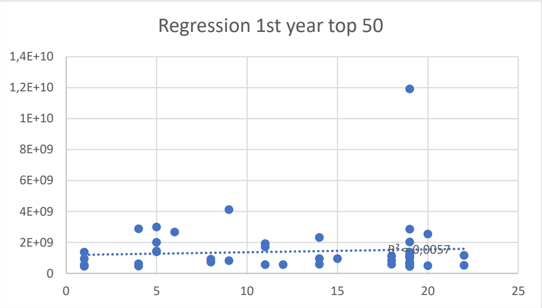
APPENDICES

Appendix 1. Top 50, lowest 50 regressions from three years, from all the years summed and their corresponding scatter plots

Table 1: Regression output for 1st year top 50 volumes

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0,075787
R Square	0,0057437
Adjusted R Square	-0,01497
Standard Error	1,743E+09
Observations	50



ANOVA

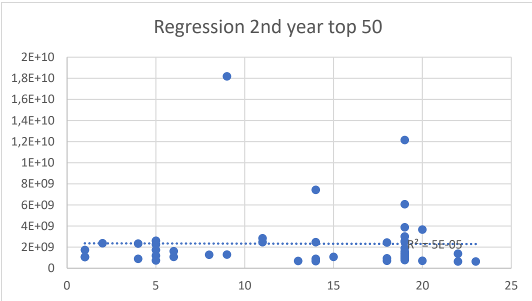
	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	8,42195E+17	8E+17	0,2773	0,60090903
Residual	48	1,45788E+20	3E+18		
Total	49	1,4663E+20			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95,0%*	*Upper 95,0%*
Intercept	1,181E+09	501599987,4	2,3536	0,0227	172007938	2,189E+09	172007938	2,189E+09
X Variable 1	18374530	34893964,93	0,5266	0,6009	-51784489	88533548	-51784489	88533548

Source: Author’s own calculations

Table 2: Regression output for 2nd year top 50 volumes

SUMMARY OUTPUT



Regression Statistics	
Multiple R	0,0067463
R Square	0,0000455
Adjusted R Square	-0,0207869
Standard Error	3046970844
Observations	50

ANOVA

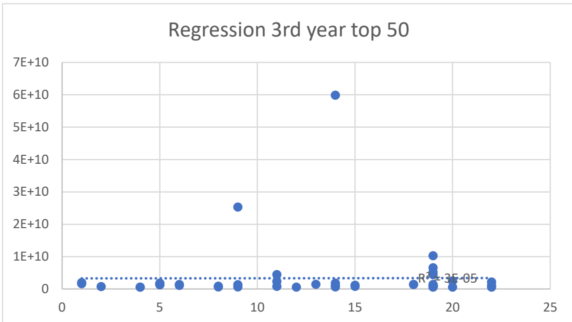
	df	SS	MS	F	Significance F
Regression	1	2,02829E+16	2E+16	0,0022	0,96291374
Residual	48	4,45634E+20	9E+18		
Total	49	4,45654E+20			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95,0%	Upper 95,0%
Intercept	2364824496	977330509,8	2,4197	0,0194	399769803	4,33E+09	399769803	4,33E+09
X Variable 1	-2992804,4	64029798,64	-0,0467	0,9629	-131733343	125747734	-131733343	125747734

Source: Author's own calculations

Table 3: Regression output for 3rd year top 50 volumes

SUMMARY OUTPUT



Regression Statistics	
Multiple R	0,0054215
R Square	0,0000294
Adjusted R Square	-0,0208033
Standard Error	9085451299
Observations	50

ANOVA

	df	SS	MS	F	Significance F
Regression	1	1,16463E+17	1E+17	0,0014	0,97019284
Residual	48	3,96218E+21	8E+19		
Total	49	3,9623E+21			

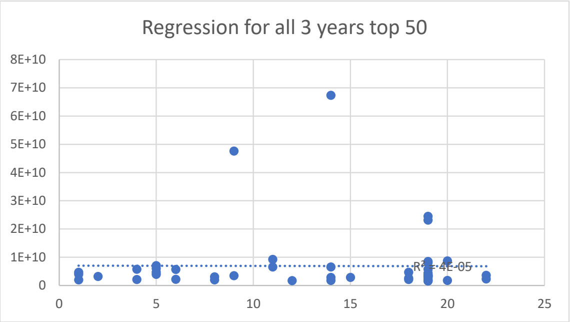
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95,0%	Upper 95,0%
Intercept	3210363516	3035772110	1,0575	0,2956	-2893465404	9,314E+09	-2,893E+09	9,314E+09
X Variable 1	7630140,69	203135539	0,0376	0,9702	-400801235	416061516	-400801235	416061516

Source: Author's own calculations

Table 4: Regression output for all years' top 50 volumes

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0,00649792
R Square	0,00004222
Adjusted R Square	-0,0207902
Standard Error	1,1629E+10
Observations	50



ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	2,74086E+17	3E+17	0,002	0,96427828
Residual	48	6,49114E+21	1E+20		
Total	49	6,49141E+21			

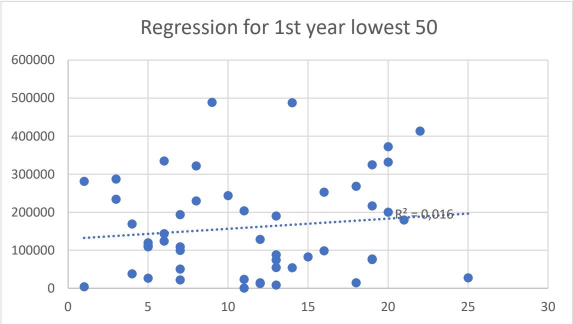
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95,0%*	*Upper 95,0%*
Intercept	7043494150	3789294496	1,8588	0,0692	-575393070	1,466E+10	-575393070	1,466E+10
X Variable 1	-11251041	249913014,7	-0,045	0,9643	-513734835	491232753	-513734835	491232753

Source: Author's own calculations

Table 5: Regression output for 1st year lowest 50 volumes

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0,12632384
R Square	0,01595771
Adjusted R Square	-0,0045432
Standard Error	130181,123
Observations	50



ANOVA

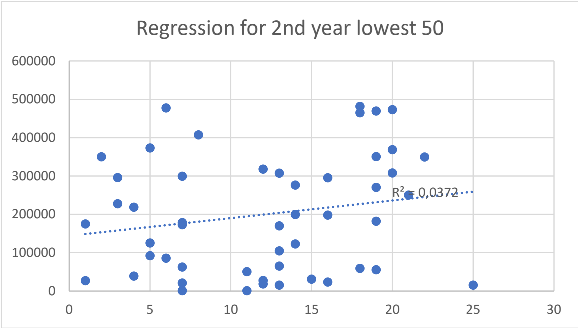
	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	13191497815	1E+10	0,7784	0,38203059
Residual	48	8,13462E+11	2E+10		
Total	49	8,26653E+11			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95,0%*	*Upper 95,0%*
Intercept	129996,505	39199,09422	3,3163	0,0017	51181,4435	208811,57	51181,4435	208811,57
X Variable 1	2664,25089	3019,785387	0,8823	0,382	-3407,43457	8735,9364	-3407,4346	8735,9364

Source: Author's own calculations

Table 6: Regression output for 2nd year lowest 50 volumes

SUMMARY OUTPUT



Regression Statistics	
Multiple R	0,192797
R Square	0,03717068
Adjusted R Square	0,01711174
Standard Error	150559,887
Observations	50

ANOVA

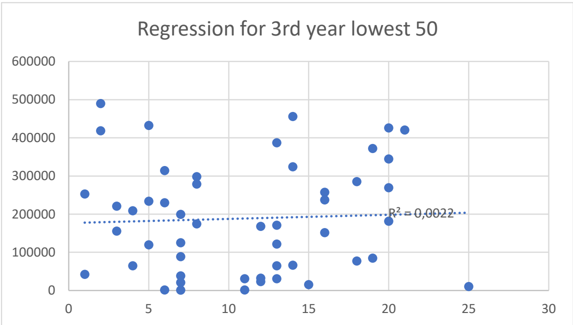
	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	42005970289	4E+10	1,8531	0,1797829
Residual	48	1,08808E+12	2E+10		
Total	49	1,13008E+12			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95,0%*	*Upper 95,0%*
Intercept	143929,993	45774,81331	3,1443	0,0029	51893,562	235966,42	51893,562	235966,42
X Variable 1	4596,71227	3376,766967	1,3613	0,1798	-2192,73276	11386,157	-2192,7328	11386,157

Source: Author’s own calculations

Table 7: Regression output for 3rd year lowest 50 volumes

SUMMARY OUTPUT



Regression Statistics	
Multiple R	0,04725279
R Square	0,00223283
Adjusted R Square	-0,018554
Standard Error	144504,499
Observations	50

ANOVA

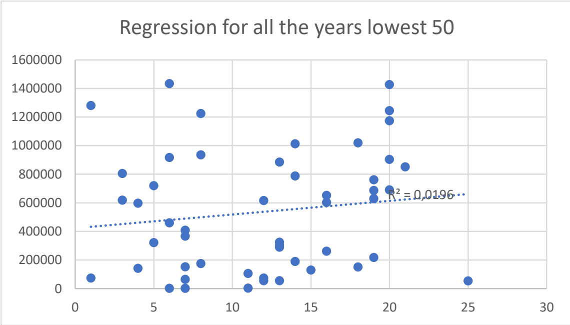
	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	2243002162	2E+09	0,1074	0,74453123
Residual	48	1,00231E+12	2E+10		
Total	49	1,00456E+12			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95,0%*	*Upper 95,0%*
Intercept	176507,129	41789,97515	4,2237	0,0001	92482,7521	260531,51	92482,7521	260531,51
X Variable 1	1092,04675	3332,021166	0,3277	0,7445	-5607,43082	7791,5243	-5607,4308	7791,5243

Source: Author’s own calculations

Table 8: Regression output for all years' lowest 50 volumes

SUMMARY OUTPUT



Regression Statistics	
Multiple R	0,13990446
R Square	0,01957326
Adjusted R Square	-0,0008523
Standard Error	420313,569
Observations	50

ANOVA

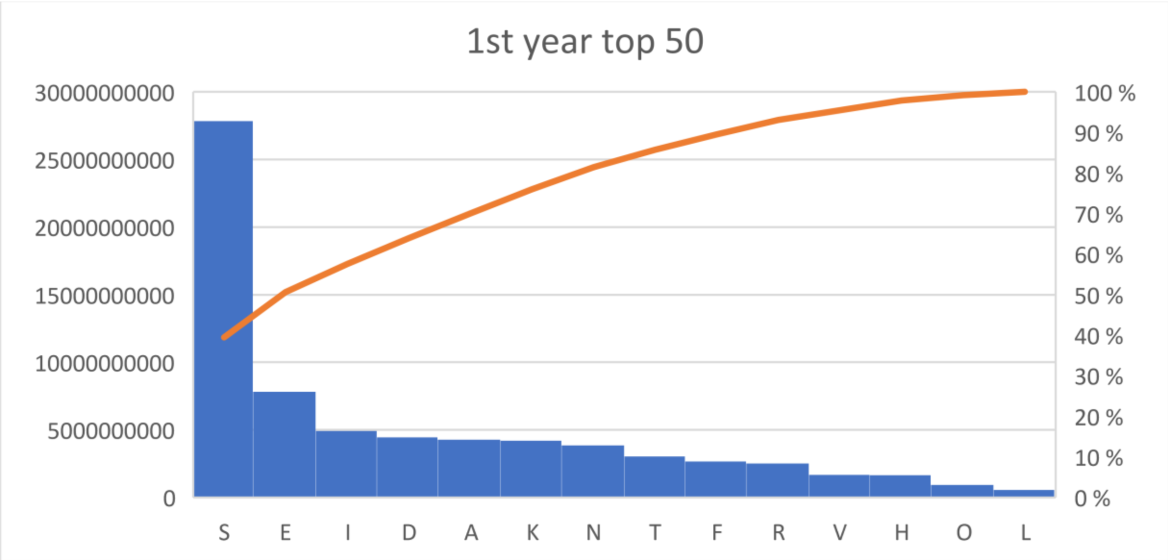
	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	1,69292E+11	2E+11	0,9583	0,33253064
Residual	48	8,47985E+12	2E+11		
Total	49	8,64914E+12			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95,0%*	*Upper 95,0%*
Intercept	423159,91	130306,6617	3,2474	0,0021	161160,807	685159,01	161160,807	685159,01
X Variable 1	9539,00752	9744,477865	0,9789	0,3325	-10053,5784	29131,593	-10053,578	29131,593

Source: Author’s own calculations

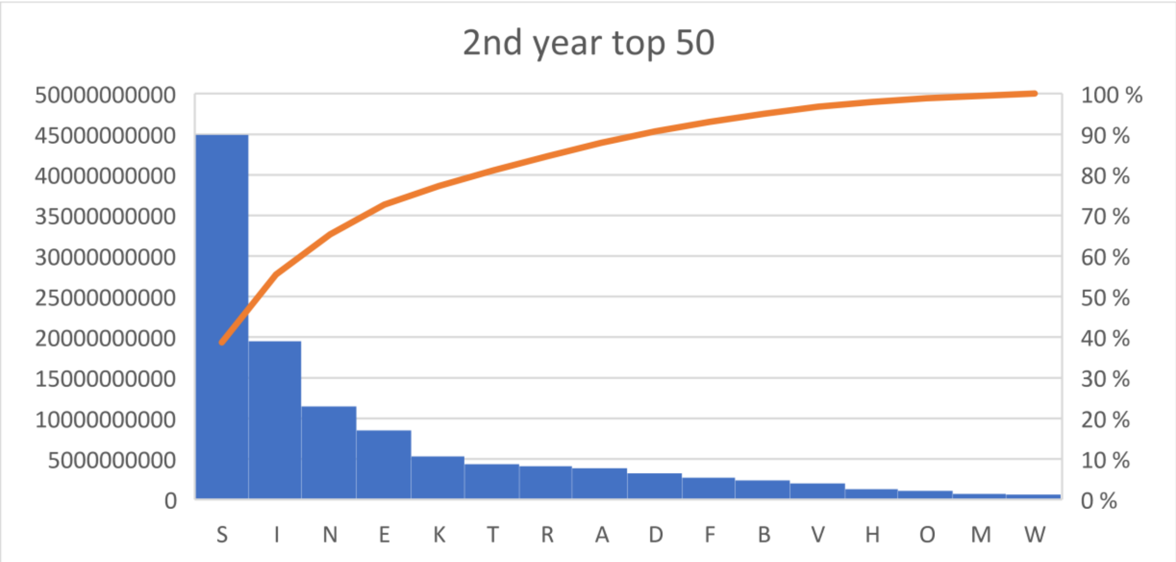
Appendix 2. Pareto charts for top 50, lowest 50 from three years and from all the years

Figure 6: Pareto chart from 1st year top 50 volume and the first letter



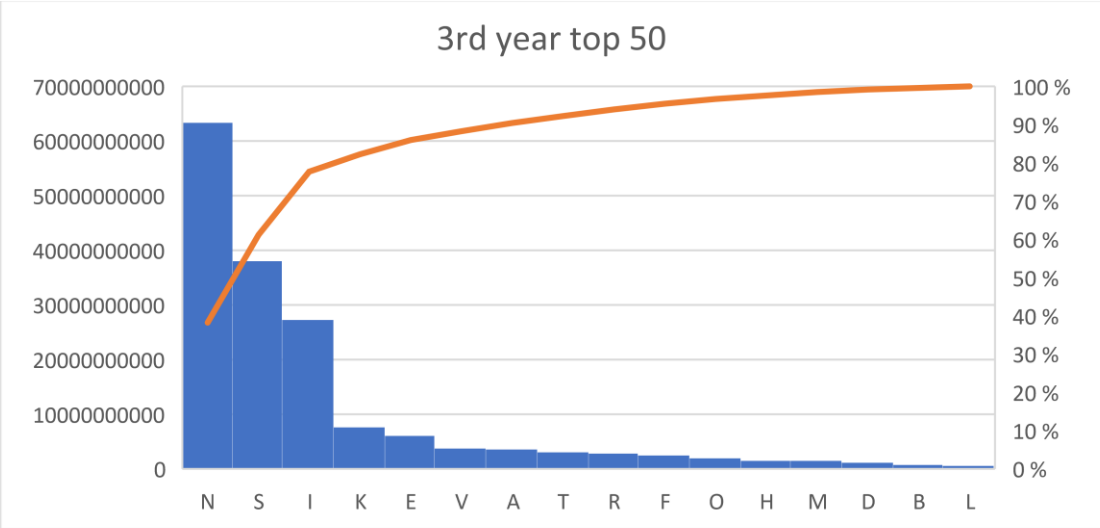
Source: Author’s own calculations

Figure 7: Pareto chart from 2nd year top 50 volume and the first letter



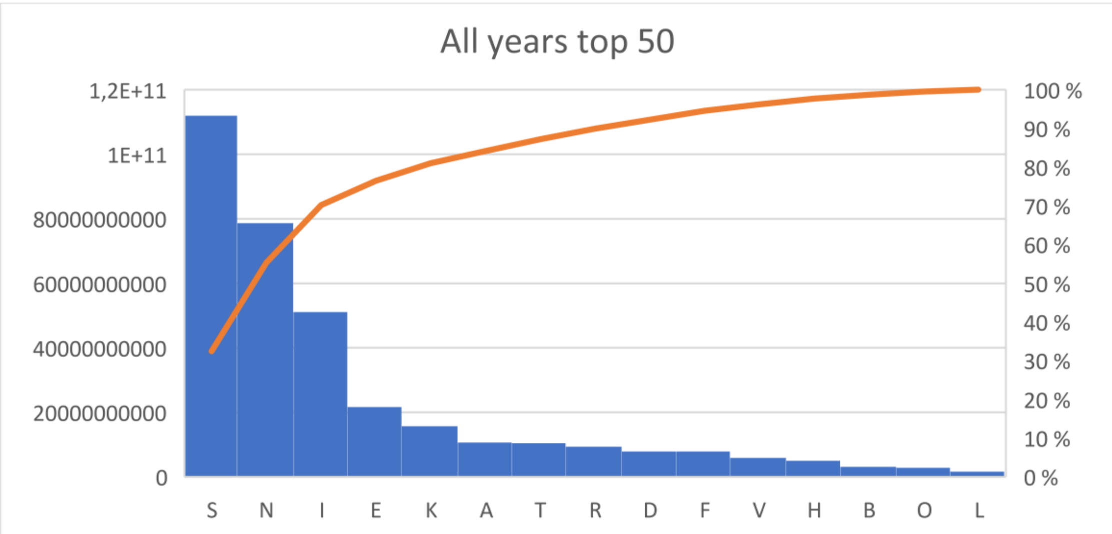
Source: Author’s own calculations

Figure 8: Pareto chart from 3rd year top 50 volume and the first letter



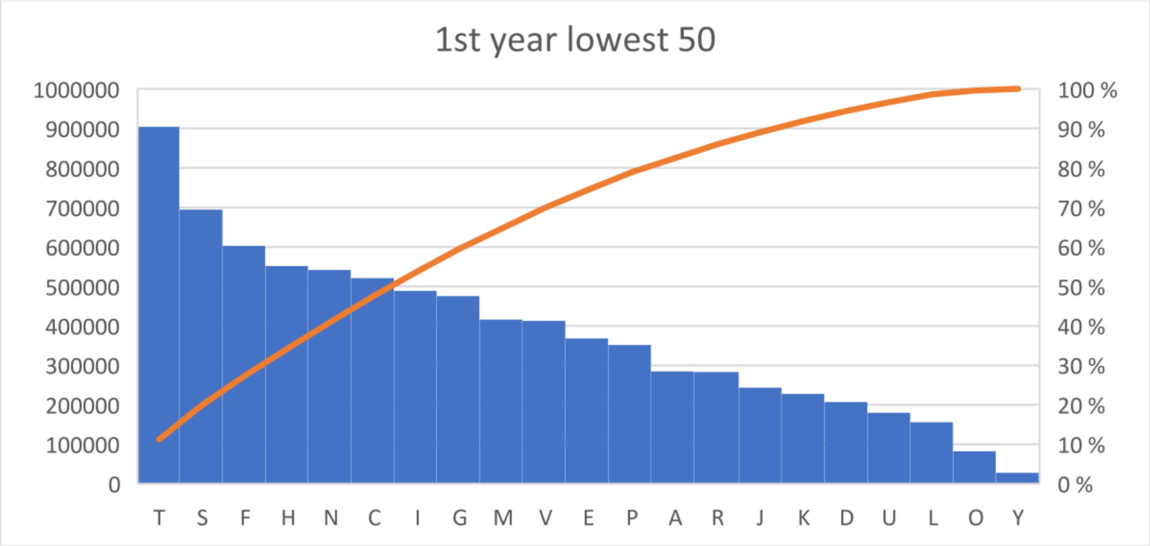
Source: Authors's own calculations

Figure 9: Pareto chart from all years' top 50 volume and the first letter



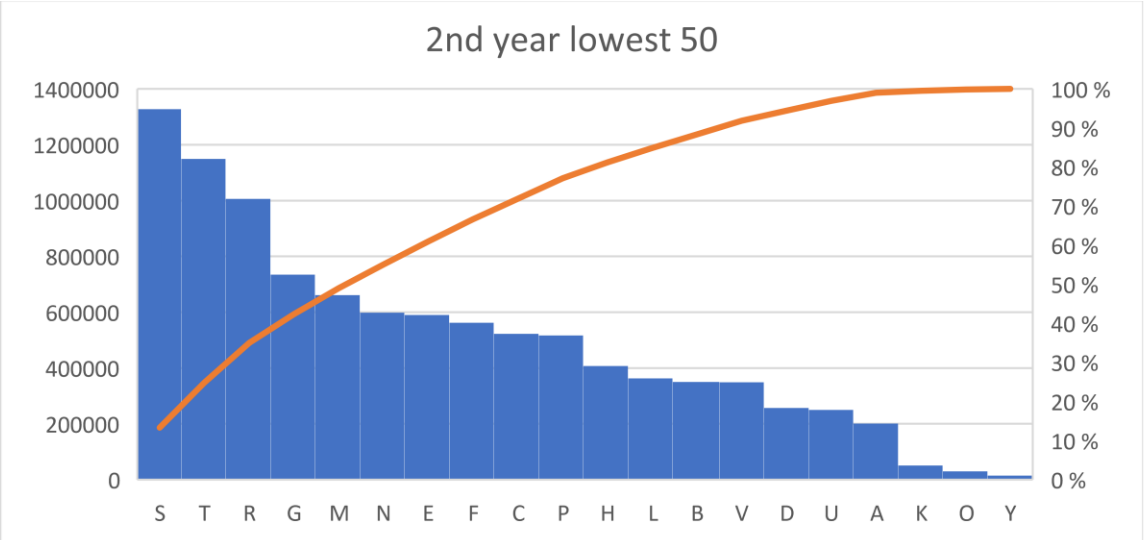
Source: Author's own calculations

Figure 10:Pareto chart from 1st year lowest 50 volume and the first letter



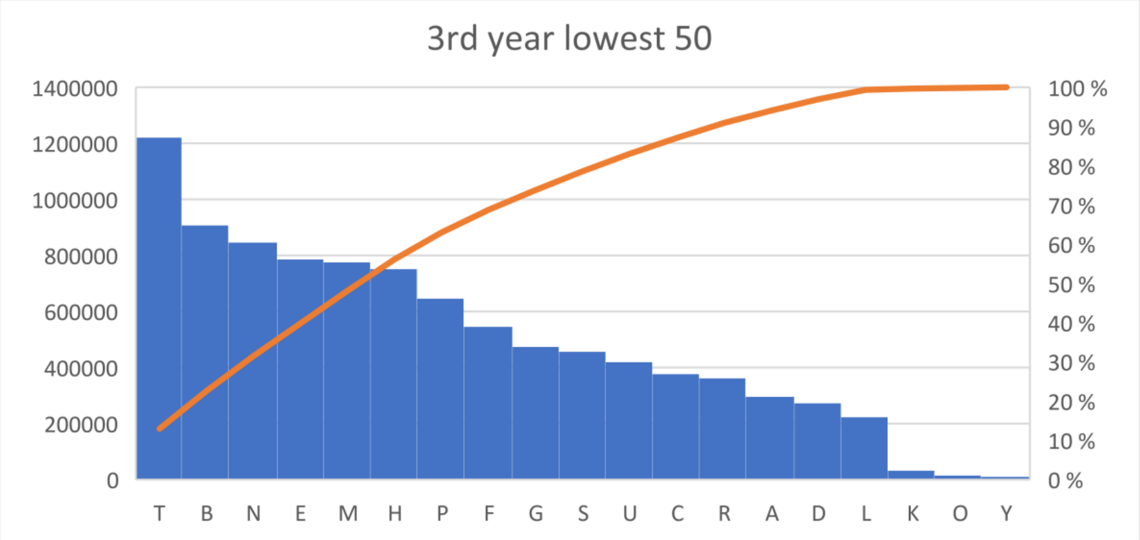
Source: Author’s own calculations

Figure 11: Pareto chart from 2nd year lowest 50 volume and the first letter



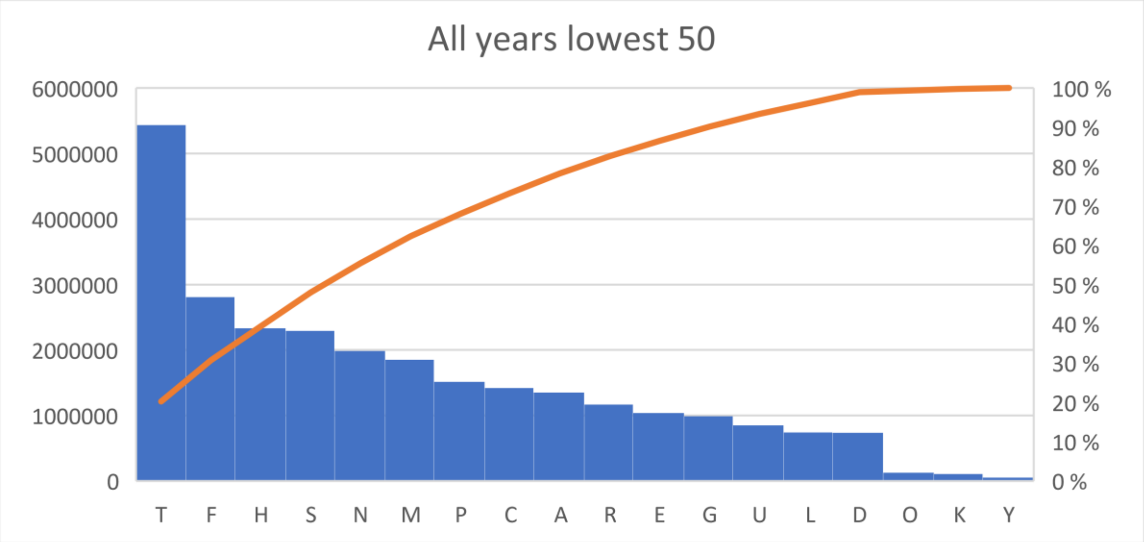
Source: Author’s own calculations

Figure 12: Pareto chart from 3rd year lowest 50 volume and the first letter



Source: Author’s own calculations

Figure 13: Pareto chart from all year’s lowest 50 volume and the first letter



Source: Author’s own calculations

Appendix 3. Link to Excel and Google Sheet files

https://1drv.ms/x/s!AkBX6wtoieCrqUeYYw4okQUdDE-W?e=IgoMvH

https://1drv.ms/x/s!AkBX6wtoieCrqUwOv7jUd5CNTRGA?e=Seaek9

https://1drv.ms/x/s!AkBX6wtoieCrqUCS0AUnXsnf472n?e=xq1T0d

https://1drv.ms/x/s!AkBX6wtoieCrqT4mejmS92PQku9S?e=g4K6Li

https://docs.google.com/spreadsheets/d/e/2PACX-
1vSzjW4502COFOS9BxAirYcfxEW0hmMrri3GKisBd9CaOW4Z-
RLMwnbp7_xyXBnllYGM98RbBVOjyd-S/pubhtml

https://docs.google.com/spreadsheets/d/e/2PACX-1vTMRX58ZqB-
sMIycaR16chhDhuEaucYpEEhhxIAOFSpW9GIEcDStrC26ll8DY5a2vitD_dXxVykm66W/pubh
tml

Appendix 4. Non-exclusive licence

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

I Leo Estola

1. Grant Tallinn University of Technology free licence (non-exclusive licence) for my thesis Alphabetic Bias in Nordic Stocks

supervised by Pavlo Illiashenko,

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.

---

[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.*