Silas Aune

Investors on social trading platforms

Empirical analysis of the disposition effect on the social trading platform Shareville

Master's thesis in Master of Science in Financial Economics
Supervisor: Jacopo Magnani
June 2023

Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Economics and Management
Department of Economics

NTNU
Norwegian University of
Science and Technology

Silas Aune

Investors on social trading platforms

Empirical analysis of the disposition effect on the social trading platform Shareville

Master's thesis in Master of Science in Financial Economics
Supervisor: Jacopo Magnani
June 2023

Norwegian University of Science and Technology
Faculty of Economics and Management
Department of Economics

NTNU
Norwegian University of Science and Technology

Preface

This master's thesis marks the end of my 5-year study at NTNU at the Department of economics. At times working with the paper was tiresome, but in the end, it was one of the most educational periods in my studies. I wish to thank my supervisor Jacopo Magnani, who introduced me to the subject and for his guidance through these months. Furthermore, I would like to give my gratitude to my friends and family for kind words and support. Finally, I would like to mention Mina, who was the deciding factor. Had it not been for her, I would not have started or completed my master's thesis.

This master's thesis is my own entirely. Sources and referrals are in the bibliography. All errors and shortcomings are entirely my own.

Silas Aune

01.06.2023

Abstract

This research paper is an empirical study of the disposition effect. The disposition effect is a phenomenon within behavioural finance. It is defined by the tendency to prematurely sell winning assets and holding on to losing assets. The analysed portfolios are collected from investors located on the social trading platform Shareville. The analysed period is from 01.01.2022 to 31.12.2022. The purpose of the research paper is to determine whether the disposition effect is present amongst the Shareville investors. Analysis is preformed to outline the predicators of the disposition effect.

In my analysis I find the presence of the disposition effect amongst the investors one the social trading platform Shareville. I find significant predicators of the disposition effect both with regards to the holding time and without regard to the holding time of the stock. The main finding is that if the stock has positive return, it is far more likely to be sold by the investor. I am not able to find significant result regarding the disposition effect varying from month to month.

Sammendrag

Denne masteroppgaven er en empirisk studie av disposisjonseffekten. Disposisjonseffekten er et velkjent fenomen innen adferdsfinans. Det defineres som tendensen til å realisere gevinster for tidlig og holde for lenge på tap. De analyserte porteføljene er hentet fra investorene på den sosiale investorplattformen Shareville. Den analyserte perioden er fra 01.01.2022 til 31.12.2022. Hensikten med masteroppgaven er å avgjøre om disposisjonseffekten er til stede blant investorene på Shareville. Ytterligere analyse er gjennomført for å finne ut hvilke forhold som påvirker disposisjonseffekten.

I min analyse finner jeg disposisjonseffekten blant investorene på den sosiale investorplattformen Shareville. Jeg finner signifikante predikatorer for disposisjonseffekten både med hensyn til hvor lenge aksjen er holdt og uten hensyn til hvor lenge aksjen er holdt. Jeg finner at hvis aksjen har positiv avkastning er det langt høyere sannsynlighet for at aksjen blir solgt av investoren. Jeg finner ikke signifikante resultater angående variasjon i disposisjonseffekten fra måned til måned.

Table of contents

Figures

Tables

1 Introduction

The disposition effect is an anomaly discovered in behavioural finance. It refers to the tendency to hold on to the assets that are losing money and prematurely sell the assets that have made financial gain (Shefrin & Statman, 1985). The paper main purpose is to analyse investors on Shareville with regards to the disposition effect. The analysed investors are the most followed investors and the experts on the social trading platform Shareville. Furthermore, the goal is to analyse and enlighten factors that influence selling decisions.

I have always been curious about behavioural finance. My motivation behind this study is to analyse if the disposition effect amongst the investors on Shareville. Furthermore, I wish to obtain a better understanding of the effect and which elements effect the most.

Social trading platforms allow investors to interact and connect with each other. In many ways these trading platforms may resemble something in the likes of Twitter or Facebook. They are like social networks that allow individual investors to share their thoughts and meanings, but most importantly their asset purchases and sales with corresponding portfolios. Most trading platforms allow users to trade or buy shares of companies, commodities or exchange trade-funds and interact with each other. Social trading platforms enable communities of investors to trade together, and one of these social trading platforms is called Shareville.

Shareville is a service that Nordnet AB offers. Nordnet Bank AB, commonly shorted to Nordnet (Nordnet, 2023), is a Nordic financial service company, headquartered in Stockholm, Sweden. When founded in 1996 it became the first internet broker in Sweden. Since then, Nordnet has expanded its services to include savings, investment, loans, and pension services. In 2011 Fabian Grapengiesser (Nordnet, 2013) founded the social trading platform Shareville. Nordnet bought the majority of Shareville in 2013 and launched the social trading platform in 2014 (Nordnet, 2014).

Shareville is Nordnets social network for saving and investing. Shareville, like other social trading platforms allow individual investors to interact and connect with each other. It also allows members to check each other’s trading. Members can see trading history, current portfolio, returns, social post, followers and much more.

2 Theoretic framework and previous studies

This chapter will highlight certain papers and previous studies related to the disposition effect. It will show how the effect was first discovered and in turn highlight its development and important finding throughout the years. The disposition effect is one of the most widely documented behavioural biases. The disposition effect is an anomaly discovered in behavioural finance. In short it relates to investors tendency to sell the winners and hold on to the losers. Starting with the discovery of prospect theory by Kahneman an Tversky, Shefrin and Statman identified the disposition effect. From there several researchers have documented the effect.

2.1 Kahneman and Tversky 1979

Research has traced the origin of the disposition effect to the so-called "prospect theory". This theory was first identified in 1979 by Daniel Kahneman and Amos Tversky. Since this paper is the origin of what we today call the disposition effect, it is a natural choice to include it in the thesis. In Kahneman and Tversky's study participant were presented with two situations. In situation 1 they have \$1000 and in situation 2 they have \$2000. In both situations they had to choose between option A and option B (Kahneman & Tversky, 1979).

Situation 1	**Situation 2**
Option A: 50% chance of gaining \$1000 and 50% chance of gaining \$0.	Option A: 50% chance of losing \$1000 and 50% chance of losing \$0.
Option B: 100% chance of gaining \$500	Option B: 100% chance of losing \$500.

An overwhelming number of participants chose option B in the first situation and option A in the second situation. This suggests that participants are willing to settle for a reasonable amount of profit (despite the decent opportunity of gaining more). However, if the risk could result in loss reduction, the study suggests that participants are far more willing to engage in risk-seeking activities.

Considering making an investment, the investor applies a personal value function, where the investor applies personal utility of the outcome and their personal belief to choose the action. The value function is defined as,

$$\boldsymbol{V(x,p;y,q)) = \pi(p)v(x) + \pi(q)v(y)}$$

where $\boldsymbol{\pi(p)}$ is defined as the decision weight for probability $\boldsymbol{p}$, and $\boldsymbol{v(x)}$ is the subjective value of the outcome $\boldsymbol{x}$. The last part of the right-hand side of the equations is interpreted equally in the same way for $\boldsymbol{y}$. The main claim in Kahneman and Tversky study is that here is evidence for $\boldsymbol{0 < p < 1, \pi(p) + \pi(1-p) < p}$. In their paper they named this property sub certainty.

From this Kahneman and Tversky defined the value function as a deviation from the reference point. The reference point would be the initial value of the investment made. The most important feature of the value function is than it is convex for losses and concave for gains. In other words, the investors are risk-averse for gains (above the reference point,) and risk-lovers for losses (below the reference point).

$$\boldsymbol{\frac{\delta V(x,p;y,q)}{\delta x} < for\ x > 0} \qquad \boldsymbol{\frac{\delta V(x,p;y,q)}{\delta x} > for\ x < 0}$$

In this sense, the participants value their losses more than they do the same amount of gains. This phenomenon is called the "Asymmetric value function". Which means, in short, the perceived pain of loss outweighs the equivalent level of gain. The asymmetric value function is displayed in figure 1.

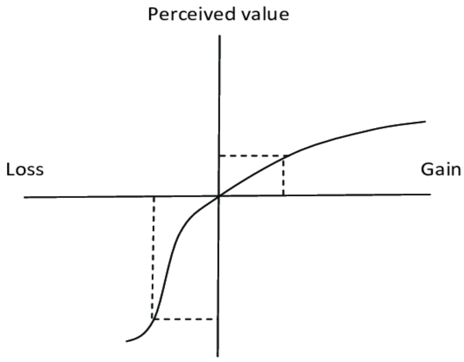

Figure 1: Asymetric value function

2.2 Shefrin and Statman 1985

The disposition effect was identified and named by Hersh Shefrin and Meir Statman in 1985. In their paper they examine the decisions to realize gains and losses in a market setting. Specifically, they focused on financial markets and whether investors exhibit a reluctance to realize losses (disposition to "ride losers") even when the precepts of standard theory prescribe realization. Shefrin and Statman noted that individuals do not like causing losses any more than they like making benefits. They also noted that individuals are far more willing to gamble on losses. Consequently, investors had a disposition to sell stocks that have risen in value but holding on to stocks that have decreased in value. Shefrin and Statman named the term "disposition effect" to describe the tendency of holding on to the loser-stocks and selling of the winner-stocks. It is worth noting that Shefrin and Statman did not prove or conclude the disposition effect in their paper. What they did conclude was that tax considerations cannot explain the pattern of loss and gain realizations itself, but their finding motivated other to continue working on the disposition effect (Shefrin & Statman, 1985).

2.3 Odean 1998

Terrance Odean developed a method to measure the disposition effect. A method used in several later studies. In Odean's method there are four ways to categorize stocks once a selling decision is made.

$$\frac{\boldsymbol{Realized\ Gains}}{\boldsymbol{Realized\ Gains + Paper\ Gains}} = \boldsymbol{Proportion\ of\ Gains\ Realized\ (PGR)}$$

$$\frac{\boldsymbol{Realized\ Losses}}{\boldsymbol{Realized\ Losses + Paper\ Losses}} = \boldsymbol{Proportion\ of\ Losses\ realized\ (PLR)}$$

If these two equations indicate significant differences in PGR and PLR then investors are, on average, more willing to realize either gains or losses. In the case of the disposition effect, PGR is higher than PLR. In Odean's study of 10 000 accounts in a U.S brokerage from 1987 through 1993. He discovered that on average 14,8% of the gains available for realization were realized (PGR), while only 9,8% of the losses were realized (PLR). Thus, investors are 50% more likely to realize gains than losses (Odean, 1998).

2.4 Grinblatt and Keloharju 2001

Another important study on the disposition effect is the 2001 study by Grinblatt and Keloharju. The unique dataset allowed them to monitor buys, sells and holds of individuals and institutions in the Finnish stock market. The dataset is a register of the official daily recording, from December 27, 1994, through January 10, 1997, of the shareholdings and trades of virtually all Finnish investors – both retail and institutional. In their study, they used a logit regression method for estimating the disposition effect. There are several advantages to utilizing a regression model. One of them is that you can control for investor characteristics and market conditions. In the logit regression the dependent variable is 1 for sell and zero for hold. The independent variables in the model include control variables relating to the stock (past returns), investor (portfolio value), calendar time (dummy variables for each month), and market conditions (market returns). The dispositions effect is measured by a dummy variable taking the value of 1 for selling, and zero for not selling. The study by Grinblatt and Keloharju finds the presence of the disposition effect in investors studied. Since their dataset was so comprehensive, investors studied are households, nonfinancial corporations, government institutions, not-for-profit institutions, and financial institutions. Of all the investors mentioned, financial institutions are arguably the more sophisticated investors, but the difference in in the disposition effect among the investors were surprisingly small. They calculated that for all investor types, the odds of selling a stock are roughly half for stocks with moderate losses (less than 30%) compared to those with gains. They also found that compared to other investors, financial institutions appear, to a degree, more willing to liquidate larger losses (more than 30%) (Grinblatt & Keloharju, 2001).

2.5 Social trading

The time has come to move beyond behavioral finance to "social finance".

– ***Hirshleifer (2015)***

In the early 2000 a social aspect was introduced to the financial research. Recent literature indicates that social trading may alter investors behaviour. Hong, Kubik and Stein discovered that stock market participation is highly linked to social interaction. Recent literature by Van Rooij, Lusardi and Alessi provides convincing evidence that investors listen to their non-expert friends before making financial decision. With the introduction of the internet, social trading platforms in turn emerged. Heimer shows that investors who trade open to the public, the disposition effect tends to be higher than in a common trading setting without any social interaction between investors.

2.6 Hong, Kubik and Stein 2004

In 2001 Hong, Kubik and Stein investigated the idea that stock-market participation is influenced by social interaction. They built a simple model in which any given "social" investor finds it more attractive to invest in the market when the participation rate among his peers is higher. They used data collected from the Health and Retirement Study (HRS) administered by the Institute for Social research at the University of Michigan. The data consists of approximately 7500 households who have a member born during the period 1931-1941. The first survey was conducted in 1992 and was named "wave 1". Three consecutive studies were conducted in the years 1994, 1996 and 1998. Three significant findings that emerged from their study. First, the model predicted a higher participation rate among social investors than among "non-social" investors. Social households – defined as those who interact with their neighbours, or who attend church – are indeed substantially more likely to invest in the stock market than "non-social" households, controlling for other factors like wealth, race, education, and risk tolerance. Second, the impact of sociability is much stronger in states where stock-market participation is higher. Third, the difference between social and “non-social“ households appears to have widened over the course of the 1990’s, as the overall stock-market participation have climbed sharply. Another key finding in their study was the indication of multiple social equilibria (Hong, Kubik & Stein, 2004).

2.7 Rooij, Lusardi and Alessi 2011

Rooij, Lusardi and Alessi devised two modules for De Nederlandsche Bank (DNB) Household Survey to measure financial literacy and study the relationship to stock market participation. They used data collected from the 2005 survey conducted by De Nederlandsche Bank on household demographic and economic characteristics and focusing on wealth and saving data. The first section of their two-part module was aimed to assess basic financial literacy. Question was in the range of interest rate, inflation, discounting, nominal versus real values. The second set of questions aims to measure more advanced financial knowledge. In their study they found that the majority of respondents display basic financial knowledge and have some grasp of the basic concepts. Most importantly they found that financial literacy affects financial decision-making (Rooij, Lusardi & Alessi, 2011).

2.8 Heimer 2016

In 2016, Heimer presented study about social interaction and the disposition effect. Heimer used data collected from the social trading network myForexBook. In total 5693 investors were relevant to the study. Those investors made roughly 2.2 million trades in the period early 2009 to December 2010. After some trimming of the data, the dataset is reduced to 2598 investors. Heimer concluded that social interaction contributes to some investors’ disposition effect. To credibly estimate casual peer effects, Heimer exploited the staggered entry of retail brokerages into partnerships with the social trading platform and compared investor activity before and after their exposure to these new social conditions. Heimer discovered that access to the social network nearly doubles the magnitude of an investor’s disposition effect and that investors connected in the network develop correlated levels of the disposition effect (Heimer, 2016).

2.9 Pelster and Hofmann 2018

With similarities to Heimer, Pelster and Hofman also studied social trading with regard to the disposition effect. In Pelster and Hofmann’s article they studied the relationship between giving financial advice and the disposition effect. The fear about reputational loss. Their data consists of investors on the online trading platform eToro. The data consists of all trades that took place between January 1, 2012, and October 8, 2015. Investors who executed less than 5 trades in this period were excluded from the dataset. In total 150 million trades and 354 817 investors were included. A commonly used feature on eToro is their investors' ability to copy other platform users. This led Pelster and Hofmann to divide their data into four different groups (the percentage each group constitutes is given).

1. Manually duplicate the investment strategies of other investors ("Advice investors"), 5,0%.
2. Automatically duplicate the investment strategy of one or more selected other investors ("Delegation investors"), 71,7%.
3. Have their investment strategies copied by other investors ("Leader investors"), 4,2%
4. Execute trades on their own without copying other investors or being copied ("Autonomous investors"), 19,1%.

Their difference-in-difference analysis suggest a significant correlation between the number of investors copying trading strategies of others and the manifestation of their disposition effect. Investors with many followers, the leader treaders, are less likely to close losing positions. Investors who just received the title "Leader investor", significantly increase their disposition effect in response to this event. Delegation investors and Advice investors do not seem to be particularly attracted to the disposition effect, this in turn provides evidence that the findings are not the result of a selection bias. Another interesting finding in Pelster and Hofmann's study is that regardless of the number of followers, female investors are always more susceptible to the disposition effect than male investors. Pelster and Hofmann come with a concluding suggestion that the disposition effect is boosted by fear of reputational loss when Leader investors are observed by their peers (Pelster & Hofmann, 2018).

3 Data material

In this chapter, I will explain how the data was collected. An explanation of the variables will be presented. Furthermore, the categorisation of the variables is presented. An explanation on how the data was processed in Excel will be presented. The data processing part will also be discussed. The complete variable list with definitions is in the appendix.

The data material analysed in this paper is manually collected from Shareville (Shareville, 2023). The data is stored, categorized and processed in Excel. The data collected from the Shareville investors are:

1. Name of stock.
2. Date of selling/purchasing.
3. Price when selling/purchasing.
4. Currency
5. Name of member.
6. Follower count.

Shareville has around 450 000 different portfolios to choose from. The 30 portfolios that are analysed in this paper is manually selected form pre-set categories on Shareville. The categories are *Experts* and *Most popular*. Within the *Experts* category we find the *experts* selected by Nordnet. There are in total 16 members in this category. Next category *Most popular* contains Shareville members with the highest follower count, note that the experts are not included in the *Most popular* category. In the final dataset there is 1 portfolio collected from the category *Experts* and 29 portfolios from the category *Most popular*. Because of the specific requirements for the data, not all portfolios are eligible for collection. From the category *Most popular*, investors are checked if they are eligible to be added to the dataset in descending order, by follower count.

The data needs to fills certain requirements. The purchasing decisions is collected from the period 01.01.2022 – 31.06.2022 and the selling decisions is collected from the period 01.07.2022 – 31.12.2022. To have sufficient data to analyse, there needs to be preferably around 10 purchasing and 5 selling decisions from each investor in the given timeframes. The sold stock needs to be one of the previously purchased stocks. In total there are 244 purchasing decisions and 109 selling decisions across the 30 different investors. The purchasing decision and selling decisions are linked to the actual date of the transaction, the price, and the currency. After collecting a sufficient amount of data, the data processing part may commence.

Through the function *stockhistory* in Excel, the daily opening price of the purchased stock, on every market day, by every investor is displayed in Excel. To make sure the stock information displayed by Excel is correct, I manually compare Excel's stock to the stock referenced on Shareville. Once the financial data is implemented into Excel, the variable collection with regards to selling decisions can commence. Every time a selling decision is made a few variables is collected.

1. Gain or loss on sold stock(s)
2. Gain or loss on held stock(s).
3. Holding time in days since purchase.

From this the data is more accurately categorized into variables. The gain or loss on sold stock(s) results in the variables RG or RL. The gain or loss on held stock(s) results in the variables PG or PL. From this, 4 variables are defined.

1. RG - Stock sold with gain.
2. RL - Stock sold with loss.
3. PG - Stock held with gain.
4. PL - Stock held with loss.

From the financial data in Excel, I can calculate the return on any given stock. This is used to measure the return on the stocks when a selling decision is made. The primary variable for this is defined as *Gain*, simply defined as weather the stock is in positive or negative return. From this the return is more precisely categorised. The return is divided into 20 different categories depending on the return. The return variables have 10% return intervals, and they cover any return from negative 100% to positive 90%. The final category of positive returns includes all observations in excess of 90%. Detailed description is attached in the appendix.

The last variable utilized in the analysis is *Holdingtime*. This is simply the number of days the stock has been held since purchase.

4 Methodology

In this chapter, I will review selected methodology used to analyse the data. Several different models and methods will be presented. They are in turn applied to the data in the next chapter. The study will focus on PGR and PGL, logistic regression and survival time analysis with and without Weibull distribution.

4.1 PGR and PGL

The PGR and PGL analysis are perhaps the simplest way to calculate the disposition effect. The method was developed by Odean in 1998. This method calculates the proportion of gains or losses realized by a given investor. A higher value in PGR than PGL would indicate that the investor realizes a higher proportion of gains than losses. The equations are given as,

$$\frac{\boldsymbol{Realized\ Gains}}{\boldsymbol{Realized\ Gains + Paper\ Gains}} = \boldsymbol{Proportion\ of\ Gains\ Realized}\ (\boldsymbol{PGR})$$

$$\frac{\boldsymbol{Realized\ Losses}}{\boldsymbol{Realized\ Losses + Paper\ Losses}} = \boldsymbol{Proportion\ of\ Losses\ realized}\ (\boldsymbol{PLR})$$

This method calculates the proportion of gains or losses realized by a given investor. These equations require 2 variables each to calculate the PGR and PLR values. The required variables are RG, RL, PG and PL. Stocks are categorized by these variables every time a selling decision is made. The different variables depend on the state of the stock. RG and RL, short for realized gain and realized loss, are stocks that are realized, or in other words sold. When sold, the stock has either positive or negative return. If the return is positive, it is classified as realized gain. If the return is negative, it is classified as realized loss. Stocks that are not sold by the investor when a selling decision is made are classified as PG and PL, short for paper gain and paper loss. These are stocks that are either in state of positive return or a state of negative return. Once these observations are collected, PGR and PLR are calculated. From PGR and PLR, the disposition effect can be calculated.

$$\boldsymbol{Disposition\ effect}: \boldsymbol{DE = PGR - PLR > 0}$$

The DE-value is a measure of the disposition effect (Odean, 1998). To test the significance of the disposition effect a statistic is calculated using the formula,

$$t = \frac{\overline{X}_1 - \mu}{\sqrt{\frac{S}{n}}} = \frac{PGR - PLR}{\sqrt{\frac{PGR - (1 - PGR)}{NRG + NPG} + \frac{PLR - (1 - PLR)}{NRL + NPL}}}$$

The NRG, NRL, NPG and NPL are the number of observations in each category. The statistic determines if the means of two populations are equal. The number gives an indication on the significance of the results (Hayes, 2023). To interpret the t-value the degrees of freedom are required. Given the sample in this research paper the formula for calculation degrees of freedom is as follows,

$$n - 1$$

Once the t-value is collected. The corresponding p-value is collected from the table in the appendix.

4.2 Logistic regression

In this part analysis the logistic regression is utilized. This method was presented in the 2001 study by Grinblatt and Keloharju. Logistic regression predicts if something is true or false. With regards to this paper that would be sold or not sold. Logistic regression has many similarities to standard linear regression. Linear regression is used when the dependable variable is measured on a continuous scale and logistic regression is used to predict a categorical dependable variable. The regression functions given as,

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n + \varepsilon$$

In the equation over y is the dependable variable, $x_1, \ldots x_n$ is a set of n independent variables, $\beta_0, \beta_1, \ldots \beta_n$ is the regression coefficients and ε is the model error. The regression coefficients $\beta_0, \ldots \beta_n$ gives the effect between the independent variables $x_1, \ldots x_n$ and the dependent variable y. The dependable variable in this paper is binary. The optimal regression model would then be logistic regression. In the logistic regression we presume that we can model p , the probability, as a function of the independent variables $x_1, \ldots x_n$. The formula for logistic regression is given as,

$$p = \frac{e^{\beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n}}{1 + e^{\beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n}}$$

The logistic regressions the purpose is to model the probability $\boldsymbol{p}$ for the dependent variable. In this case the $\boldsymbol{p}$ value will always be between 0 and 1, thereby the model now predicts the probability of the dependable variable with regards to $\boldsymbol{\beta_0}, \dots \boldsymbol{\beta_n}$ and $\boldsymbol{x_1}, \dots \boldsymbol{x_n}$. Displayed in figure 2 is the probability distribution of logistic regression. With a simple rearrangement the formula can be expressed as,

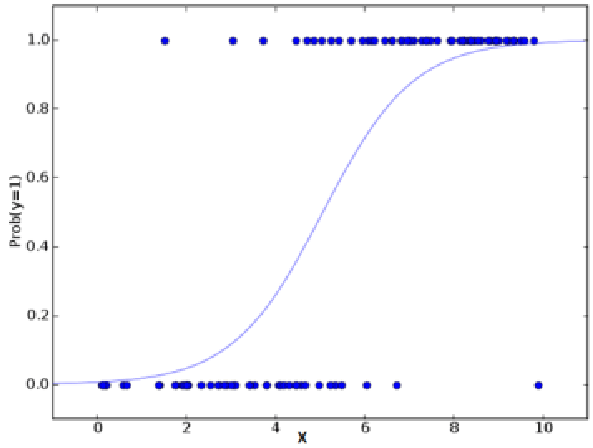

Figure 2: Logistic regression probability Distribution

$$\mathbf{ln}\left(\frac{\boldsymbol{p}}{\mathbf{1}-\boldsymbol{p}}\right) = \boldsymbol{\beta_0} + \boldsymbol{\beta_1 x_1} + \cdots + \boldsymbol{\beta_n x_n}$$

The ln-function represent the natural logarithmic function with base number $\boldsymbol{e}$. The size $\left(\frac{p}{1-p}\right)$ is called odds, and the function ln(odds) is called the logit function. The similarities with the linear regression model are now present, as the right-hand side of the regression equation is a linear function of the independent variables, but there is clear difference. In this model the regression coefficient does not give a direct link between the independent variables and response, since the left-hand side of the equation is not the dependable $\boldsymbol{y}$ variable, but $\boldsymbol{ln}\left(\frac{\boldsymbol{p}}{\mathbf{1}-\boldsymbol{p}}\right)$. Therefore, another transformation of the regression coefficients is necessary to make an interpretable result. To make this as simple as possible we include one independent variable $\boldsymbol{x}$. By using the exponential function on both sides of the equation we are left with,

$$\frac{\boldsymbol{p}}{\mathbf{1}-\boldsymbol{p}} = \boldsymbol{e}^{\boldsymbol{\beta_0}+\boldsymbol{\beta_1 x}}$$

Where $\left(\frac{\boldsymbol{p}}{\mathbf{1}-\boldsymbol{p}}\right)$ is the odds for the given response. For $\boldsymbol{x} = \mathbf{1}$ the odds become $\boldsymbol{e}^{\boldsymbol{\beta_0}+\boldsymbol{\beta_1}}$, and for $\boldsymbol{x} = \mathbf{0}$ the odds become $\boldsymbol{e}^{\boldsymbol{\beta_0}}$. The relationship between these number is known as the odds relationship, a highly used measure of effect, and we are left with,

$$\frac{\boldsymbol{e}^{\boldsymbol{\beta_0}+\boldsymbol{\beta_1 x}}}{\boldsymbol{e}^{\boldsymbol{\beta_0}}} = \boldsymbol{e}^{\boldsymbol{\beta_1}}$$

In logistic regression the measure of effect is not the regression coefficient $\boldsymbol{\beta}$, but $\boldsymbol{e}^{\boldsymbol{\beta}}$ since this gives the odds relationship (Thoresen, 2017).

4.3 Cox Proportional hazard model

Now the Cox proportional hazard model is applied to the data. The purpose of the model is to evaluate simultaneously the effect of several factors on survival. With similarity to the logistic regression, survival in this context, is the stock not being sold by a given investor. This element of the Cox proportional hazard models is advantageous because it allows for examination on how specified factors influence the rate of a particular event happening at a particular point in time. This in known as the hazard rate. The model is expressed by the hazard function denoted by $\boldsymbol{h(t)}$. In the case of this paper, the hazard function can be interpreted as the risk of selling at time $\boldsymbol{t}$. It is estimated as,

$$\boldsymbol{h(t) = h_0 t * exp\ (b_1 x_1 + b_2 x_2 + \cdots + b_n x_n)}$$

where $\boldsymbol{t}$ represent the survival time, $\boldsymbol{h(t)}$ is the hazard function determined by a set of $\boldsymbol{n}$ covariates $\boldsymbol{x_1 + x_2 + \cdots + x_n}$. The coefficients $\boldsymbol{b_1 + b_2 + \cdots + b_n}$ measure the impact of the covariates. The term $\boldsymbol{h_0}$ is called the baseline hazard. This correlates to the value of the hazard if all the $\boldsymbol{x_i}$ are equal to zero. The quantity $\boldsymbol{exp(0) = 1}$. The $\boldsymbol{t}$ in $\boldsymbol{h(t)}$ indicates that the hazard may vary over time. The $\mathbf{exp}\ (\boldsymbol{b_i})$ quantities are called hazard ratios. A value of $\boldsymbol{b_i}$ greater than zero, or equivalently a hazard ratio greater than one, indicates that as the value of the $\boldsymbol{i^n}$ covariate increases, the event hazard increases and thus the length of survival decreases. In simpler terms, a hazard ratio above 1 indicates that the given covariate is positively associated with the event probability, and a hazard ratio below 1 is negatively associated with the event probability,

- HR = 1: No effect.
- HR < 1: Reduction in the hazard.
- HR > 1: Increase in the hazard.

One of the key assumptions of the Cox proportional hazard model is that the hazard curves for the groups of observations should be proportional and cannot cross. To illustrate this, consider two groups $\boldsymbol{K}$ and $\boldsymbol{K'}$ that differ in their x-values. The two hazard functions are,

Hazard function for the group $\boldsymbol{K}$	Hazard function for group $\boldsymbol{K'}$
$\boldsymbol{h_K(t) = h_0(t) e^{\sum_{t=1}^{n} \beta_x}}$	$\boldsymbol{h_{K'}(t) = h_0(t) e^{\sum_{t=1}^{n} \beta_{x'}}}$

then in turn the hazard ratio of these two groups is independent of time $\boldsymbol{t}$. The equation is as follows.

$$[\frac{h_K(t)}{h_{K'}(t)} = \frac{h_0(t)e^{\sum_{t=1}^{n}\beta_x}}{h_0(t)e^{\sum_{t=1}^{n}\beta_{x'}}} = \frac{e^{\sum_{t=1}^{n}\beta_x}}{e^{\sum_{t=1}^{n}\beta_{x'}}}]$$

As one can see from the equation over the hazard of the event in any group is a constant multiplier of the hazard in another group. This equation confirms one of the key assumptions of the Cox proportional hazard model. The key assumption is the curves of the group should be proportional and cannot cross (Cox, 1972).

4.4 Weibull distribution

The cox proportional hazard model is used in survival analysis when there is a constant change and probability of the survival function. Using what is known as a Weibull distribution on the survival time analysis, one can measure a non-constant change in the survival function. When testing for the disposition effect there is not a constant element or a probability of holding time, therefore a Weibull distribution in the survival analysis, could be more fruitful (Seasholes and Feng 2005).

When utilizing the Weibull distribution in survival analysis, one can do a regression with a non-constant change. With the Weibull distribution in survival time analysis the hazard function is of the form,

$$h(t, p, X, Z_t) = p\lambda^{p-1}\exp\,(X\beta + Z_t\gamma + \epsilon$$

Here X is a fixed covariate and time varying covariate Z_t. Using the Weibull distribution with the following equations:

$f(t) = p\lambda t^{p-1}\exp\,(-\lambda t^p)$, Duration density

$S(t) = \exp\,(-\lambda t^p)$, Distribution of survival times.

$h(t) = p\lambda t^{p-1}$, Hazard rate

In the equations above t is time p is probability and λ is the constant of integration. The corresponding hazard ratio would be,

$$hazard\ ratio(\gamma) = \frac{h(t, p, X, Z_t = 1)}{h(t, p, X, Z_t = 0)}\exp(\gamma)$$

4.5 Likelihood-Ratio Test

The likelihood ratio test is a hypothesis test that helpful when deciding between two nested models. Nested models means that one is a variant of the other. Imagine model A with variables 1, 2, 3, 4 and model B with variables 1, 2. In this case model B is nested within model A. The best model is the one that makes the data most likely. In other words, maximizes the likelihood function. The likelihood function is defined as,

$$\boldsymbol{f_n = (X - 1, \dots, X_n | \theta)}$$

The likelihood function is higher nearer the true value for $\boldsymbol{\theta}$. The test compares the two models. The null hypothesis is that the smaller model is the best.

$$\boldsymbol{LRT = -2log_e(\frac{\mathcal{L}_s(\hat{\theta})}{\mathcal{L}_g(\hat{\theta})}}$$

In the equation above test static is calculated as ratio between the log-likelihood of the simpler model (s) to the model with more parameters (g) (Glen, 2023).

4.6 Akaike's and Bayesian information criterion

Akaike's information criteria and Bayesian information criteria are two widely used methods in model selection. The different criteria are relatively similar but, there are some slight differences in interpretation. The AIC and BIC are mathematical methods used for evaluation how well a model fits the data it was generated from. They are calculated from,

1. All the independent variables included in the model.
2. The maximum likelihood estimate of the model.

When comparing the Akaike's information criteria and the Bayesian information criteria, the BIC is far more penalizing for adding additional parameters. The BIC penalizes the free parameters more strongly. The formula for calculation the two information criteria is as follows,

$$\boldsymbol{AIC = 2k - 2\ln(L)}$$

$$\boldsymbol{BIC = kln(n) - 2\ln(L)}$$

In the equations for AIC and BIC, $\boldsymbol{L}$ is the maximum value of the likelihood function, $\boldsymbol{k}$ is the number of parameters estimated by the models, and $\boldsymbol{n}$ is the number of observations (Brownlee, 2020).

5 Results

In this chapter, I will categorically go through the different models presented in the last chapter and present the results when these are applied to the dataset. Different findings made along the way will be commented. Results will be collected and in turn discussed in light of this papers research problem. The primary goal is to find evidence of the disposition effect, measure the disposition effect, find different predicators of the disposition effect, and analyse the disposition effect with regard to the time element in the dataset.

5.1 PGR and PGL

As mentioned, the PGR and PLR analysis is calculated using the variables RG, RL, PG and PL. The distribution across the 30 investors is displayed in figure 3, and the total amount in each category is listed below.

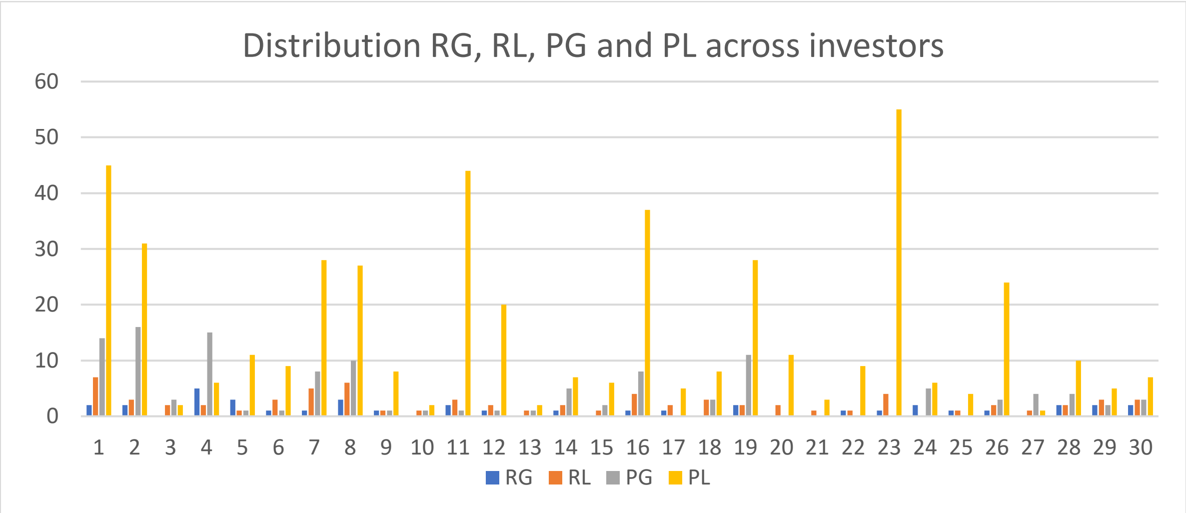

Figure 3: Distribution RG, RL, PG and PL

RG = 38 **PL = 71** **PG = 123** **PL = 461**

The corresponding PGR and PLR values are then calculated.

PGR is calculated as $$\frac{\mathbf{38}}{\mathbf{38+123}} = \mathbf{0.236}$$

PGL is calculated as $$\frac{\mathbf{71}}{\mathbf{71+461}} = \mathbf{0.133}$$

These results indicate that investors realize 23.6% of gains, but only realize 13.3% of losses. At first glance this indicates that investors are more willing to realize gains than losses.

Overall, investors realize 77,4% more gains than losses. The disposition effect can be calculated as,

$$\boldsymbol{DE = 0.236 - 0.133 = 0.103 > 0}$$

5.1.1 T-score

To validate the result a t-statistic and degrees of freedom is computed,

$$\boldsymbol{t = \frac{0.236 - 0.133}{\sqrt{\frac{0.236(1-0.236)}{38+123} + \frac{0.133(1-0.133)}{71+461}}} = 2.804}$$

From the t-value table in the Appendix the corresponding p-value is found. The t-value of 2.804 corresponds to a p-value of 0.0089. The result is significant on the 1% level, and the presence of the disposition effect is highly significant.

5.1.2 Monthly PGR and PGL

Listed below in table 1 are the monthly observations of RG, RL, PG and PL. From there table 2 has the calculated values of PGR, PGL and DE. The results are illustrated in figure 4.

	RG	RL	PG	PL
July	10	17	37	148
August	10	7	23	75
September	7	20	28	109
October	4	7	9	42
November	6	8	17	52
December	1	12	9	35

Table 1: Amount RG, RL, PG and PL by month

	PGR	PGL	DE	t-value	p-value
July	0.2128	0.1030	0.1097	1.7088	0.0981
August	0.3030	0.0854	0.2177	2.5385	0.0167
September	0.2000	0.1550	0.0450	0.6015	0.5521
October	0.3077	0.1429	0.1648	1.1995	0.2400
November	0.2609	0.1333	0.1275	1.2561	0.2191
December	0.1000	0.2553	-0.1553	-1.3599	0.1843

Table 2: PGR and PLR by month

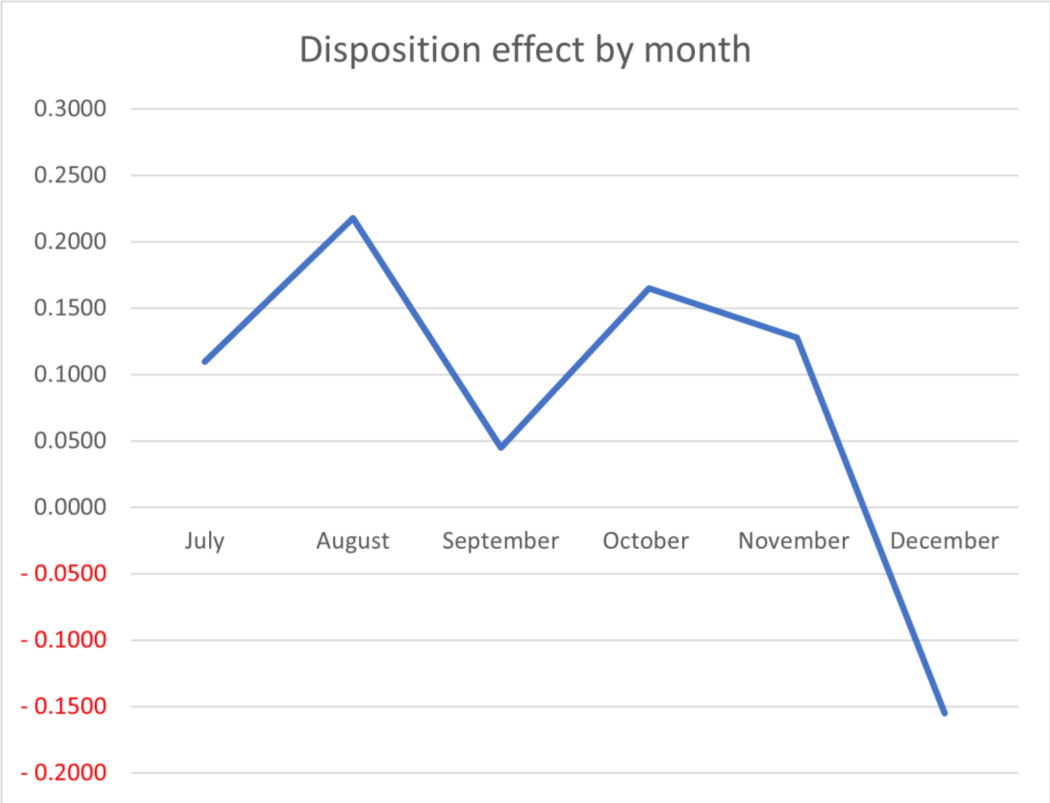

Figure 4: Disposition effect by month

As indicated by the graph in figure 4 the disposition effect is present every month except December. From November to December there is a rather large jump in the disposition effect. This is in line with previous studies indicating tax motivated selling in December (Odean, 1998), however due to few observations the result is deemed not significant. The p-value of the disposition effect in December is measured at 0.1843.

5.1.3 Monthly RG, RL, PG and PL

To better illustrate the monthly averages of RG, RL, PG and PL graphs are utilized and are displayed in figures 5-8.

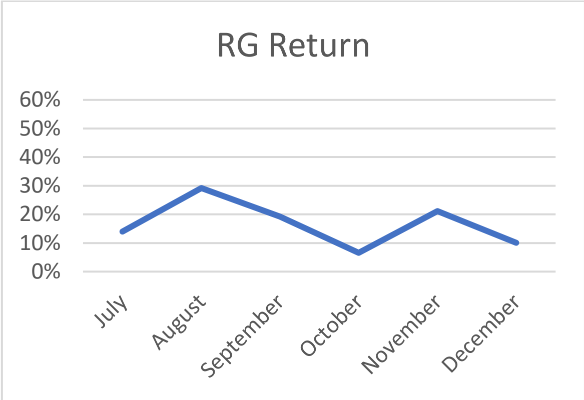

Figure 5: Monthly return average RG

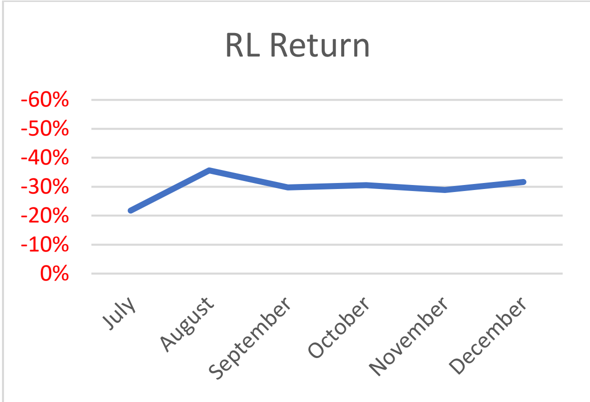

Figure 6: Monthly return average RL

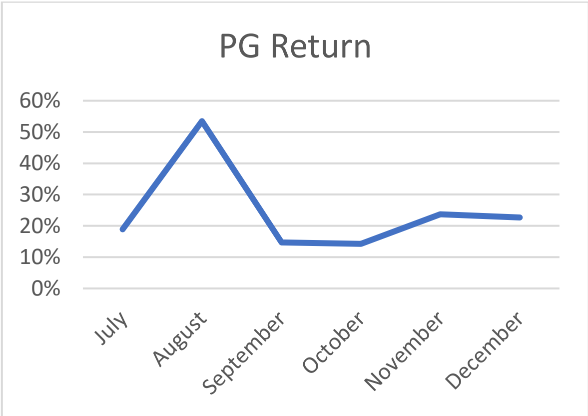

Figure 7: Monthly return average PG

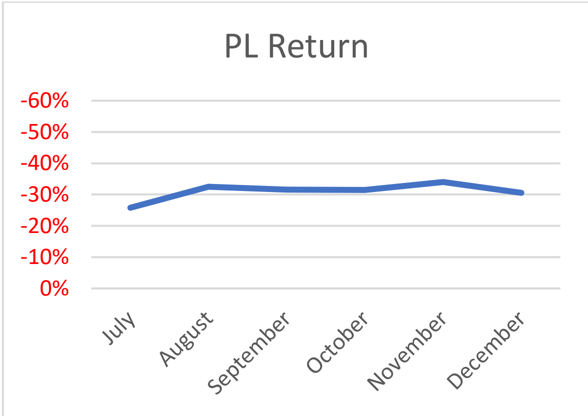

Figure 8: Monthly return average PL

As indicated by figures 5 and 7, there was a high monthly return in RG an PG in August. One can speculate that August was a profitable month in the stock market. This led to the individual month with the highest disposition effect. The investors had a huge PGR of 0.3030 and a low PGL of 0.0854. Values are collected from table 2. The calculated disposition effect in August was 0.2177. The t-value is 2.5385, corresponding to a p-value of 0.0167, meaning the result is significant at the 5% level. Through the 6-month period the RL and PL illustrated in figures 6 and 8 are observed with returns between -30% and -40%. Overall, when comparing figure 5 and 6, it looks like RL has generally a lower value than RG has a high value. This is the case for the entire 6-month period, indicating that investors let stocks get larger negative returns, than the equivalent positive returns, before they decide to sell the stock.

5.2 Logistic regression

The results in the previous chapter would indicate a presence of the disposition effect. The standard PGR and PGL method are preformed first in the analysis. To further analyse the disposition effect and to find potential predicators of the disposition effect a logistic regression is utilized. Test and model are executed in STATA. Results are collected directly from STATA.

5.2.1 Logistic regression – *Sell Gain*

Table 3 has *Sell* set as the dependent variable and *Gain* as the independent variable.

Logistic regression					Observations	693
					Chi2	9.20
					Prob > Chi2	0.0026
Log Likelihood	-297.0067				Pseudo R2	0.0151
					95% Conf. Interval	
Variable	Odds Ratio	SE	Z	P > I Z I	LL	UL
Gain	2.0059	0.4516	3.09	0.002	1.2902	3.1187
Constant	0.1540	0.0196	-14.67	0.000	0.1199	0.1977

Table 3: Logistic regression - *Sell Gain*

Table 3 includes all the 693 observations when a selling decision is made. The logistic regression model is significant at the 1% level with a Chi2 p-value of 0.0026. The independent variable *Gain* is significant at the 1% level, with a p-value of 0.002. The Log likelihood is -297.0067. The reported odds ratio for *Gain* is 2.0059. The interpretation of this is if *Gain* = 1, the chance of *Sell* = 1 is 2.0059 times as likely. Another interpretation of the odds ratio is that the *Gain* = 1 group has 100.59% higher odds of selling the stock. This is in line with the results in the PGR and PLR part of the analysis. These results are in line with previous studies using the same methodology (Grinblatt & Keloharju, 2001).

5.2.2 Logistic regression – *Sell Gain* Returns (*I, J, K, S, T* and *U*)

Table 4 includes the first 3 categories of both the gain and loss brackets. The independent variable *Gain* is still included.

Logistic regression					Observations	693
					Chi2	15.98
					Prob > Chi2	0.0253
Log Likelihood	-293,5624				Pseudo R2	0.0265
					95% Conf. Interval	
Variable	Odds Ratio	SE	Z	P > I Z I	LL	UL
Gain	2.5590	1.1700	2.06	0.040	1.0444	6.2698
I (0% to 10%)	1.0087	0.4900	0.02	0.986	0.3892	2.6141
J (10% to 20%)	0.7840	0.4364	-0.44	0.662	0.2633	2.3346
K (20% to 30%)	0.2875	0.3238	-1.11	0.268	0.0316	2.6140
S (0% to -10%)	1.7424	0.5760	1.68	0.093	0.9115	3.3308
T (-10% to -20%)	1.3226	0.4471	0.83	0.408	0.6817	2.5657
U (-20% to -30%)	0.7357	0.2968	-0.76	0.447	0.3336	1.6223
Constant	0.1359	0.0273	-9.91	0.000	0.0915	0.2017

Table 4: Logistic regression - *Sell Gain* Returns (*I*, *J*, *K*, *S*, *T* and *U*)

Number of observations is 693. The model is significant at the 5% level with a Chi2 p-value of 0.0253. Independent variables *I, J, K, S, T* and *U* in the regression model all have high p-values, they are deemed not significant. The independent variable *Gain* and dependable variable *Sell* remains significant at the 5% level, with a p-values of 0.040 and 0.000 respectively. The Log likelihood is -293.5624. There is a slight increase in the *Gain* odds ratio. The overall odds ratio for *Gain* is higher when accounting for the variables *I*, *J*, *K*, *S*, *T* and *U*. Variables *I*, *J*, *K*, *S*, *T* and *U* are binary variables, the interpretation of *I*, *J*, *K*, *S*, *T* or *U* are that if the variables are equal to 1 the corresponding odds ratio would be the increase odds for the dependable variable to be 1. For odds ratios below 1, a there is a decrease in the odds for the dependable variable to be 1. The odds ratios with the corresponding 95% confidence interval level are displayed in figure 9.

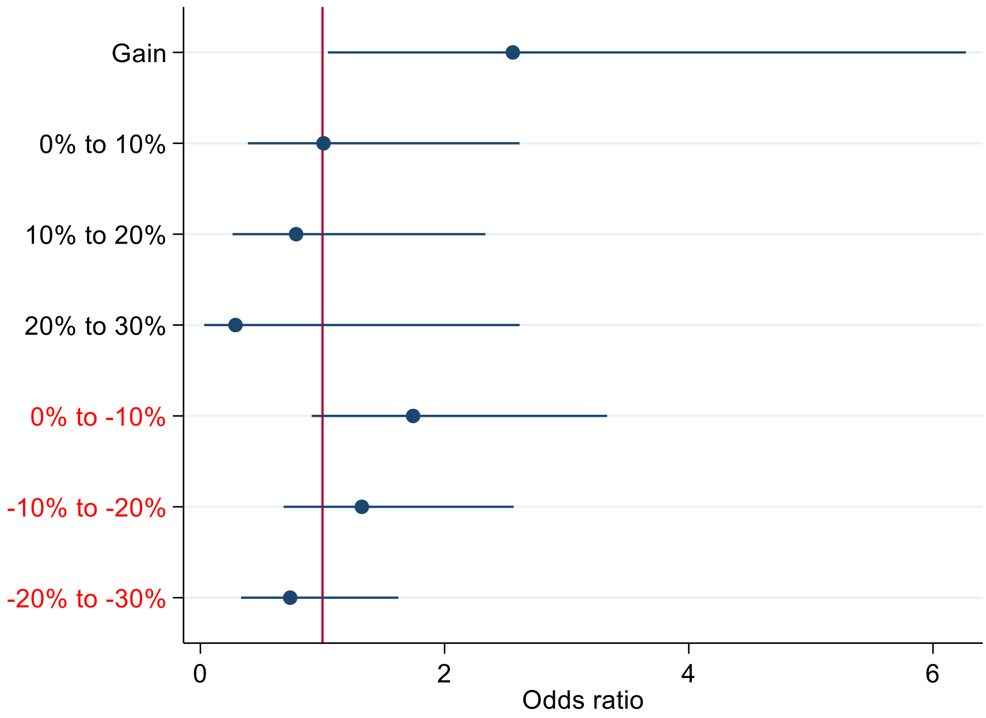

Figure 9: Odds ratios with corresponding 95% confidence interval 1

From figure 9 we can see that the negative return brackets seem to have higher odds ratios than positive return brackets, although when observing the 95% confidence interval the odds ratios are deemed not significant. This is because the each of the return brackets corresponding 95% confidence interval overlaps vertical line at odds ratio = 1. There are also some factors that needs to be addressed when including variables that explain more or less the same in the same regression model.

Including the *Gain* variable and the return variables in the same regression could be problematic since *Gain* and the positive return variables explain much of the same. From table 4 one can see that the return variables *I*, *J* and *K* indicate worse odds for the dependable variable to be 1. At first glance this would be direct opposite of what the disposition effect indicates. The reason behind this is if *I*, *J* or *K* equals 1, *Gain* would also have equal 1. This is a case of multicollinearity. Multicollinearity in short is the presence of high intercorrelations among two or more independent variables in the regression model. This can lead to skewed or misleading results (Boix, 2021). To avoid this, *Gain* will later be excluded from the regression.

5.2.3 Logistic regression – *Sell Gain* Returns (*I, J, K, S, T* and *U*) *Holdingtime*

In table 5 the variable *Holdingtime* is added. The *Holdingtime* variable is discrete, to better interpret the value of *Holdingtime,* A marginal effect analysis will later be performed.

Logistic regression					Observations	693
					Chi2	16.06
					Prob > Chi2	0.0415
Log Likelihood	-293.5242				Pseudo R2	0.0266
					95% Conf. Interval	
Variable	Odds Ratio	SE	Z	P > I Z I	LL	UL
Gain	2.5335	1.1620	2.03	0.043	1.0311	6.2249
I (0% to 10%)	0.9950	0.4859	-0.01	0.992	0.3820	2.5915
J (10% to 20%)	0.7816	0.4352	-0.44	0.658	0.2624	2.3281
K (20% to 30%)	0.2813	0.3176	-1.12	0.261	0.0307	2.5720
S (0% to -10%)	1.7117	0.5766	1.60	0.111	0.8845	3.3127
T (-10% to -20%)	1.2893	0.4518	0.73	0.468	0.6487	2.5627
U (-20% to -30%)	0.7291	0.2951	-0.78	0.435	0.3297	1.6120
Holdingtime	0.9995	0.0014	-0.28	0.783	0.9966	1.0025
Constant	0.1473	0.0520	-5.42	0.000	0.0736	0. 2946

Table 5: Logistic regression - *Sell Gain* Returns (*I, J, K, S, T* and *U*) *Holdingtime*

The number of observations remains at 693, and the overall model is still significant at the 5% level with a Chi2 p-value of 0.0415. Overall, most of the variables are not significant, except for the independent variable *Gain*. *Gain* is significant at the 5% level, with p-value of 0.043. The Log likelihood is -293.5242. The new included variable *Holdingtime* is what is known as a discrete variable. The odds ratio is 0.9995 meaning if *Holdingtime* increases 1 unit the dependable variable has its odds decreased 0.9995. To get odds ratio of a n-unit increase the percentage is required,

$$\left(\mathbf{100} * (\mathbf{1} - \mathbf{0.9995})\right) * \boldsymbol{number\ of\ units} = \boldsymbol{odds}$$

The odds ratio for *Holdingtime* is not surprisingly insignificant with a p-value of 0.783. A reason for this is length of the variable, spanning from day 10 to day 360. The cox proportional hazard model would be a more suitable model for the *Holdingtime* variable, it will be performed later in the analysis.

5.2.4 Logistic regression – Investor-fixed effects

In table 6 a dummy for each investor is added. This is to include the investor-fixed effects. This will also slightly change the interpretation of the *Gain* odds ratio.

Logistic regression					Observations	693
					Chi2	34.33
					Prob > Chi2	0.2678
Log Likelihood	-284.3891				Pseudo R2	0.0569
					95% Conf. Interval	
Variable	Odds Ratio	SE	Z	P > I Z I	LL	UL
Gain	1.8754	0.4710	2.50	0.012	1.1463	3.0683
Constant	0.1280	0.0472	-5.57	0.000	0.0621	0.2637

Table 6: Logistic model, Investor-fixed effects (Complete table displayed in appendix, table 21)

Table 6 includes all 693 observations. The Independent variable *Gain* and the dependent variable *Sell* are both significant at the 5% level, with p-values of 0.012 and 0.000 respectively. Table 6 is included to illustrate the effect on the independent variable *Gain* using only the within-investor variation. Using only the within-investor variation the odds ratio for *Gain* is 1.8754. This would be equal to treating the data as panel data and it provides a more robust answer, but the overall significance of the model is rather low with a p-value of 0.2678. The complete table including all 30 investors is displayed in the appendix.

5.2.5 Logistic *Sell* Returns (*I, J, K, S, T* and *U*)

In table 3, 4 and 5 both the *Gain* variable and the return brackets are included in the logistic regression. This could be problematic as *Gain* explain much of the same as the positive return brackets *I*, *J* and *K*. To better get the effects of the return brackets, *Gain* is excluded from this model. The results are displayed in table 7.

Logistic regression					Observations	693
					Chi2	12.19
					Prob > Chi2	0.0578
Log Likelihood	-295.4592				Pseudo R2	0.0202
					95% Conf. Interval	
Variable	Odds Ratio	SE	Z	P > I Z I	LL	UL
I (0% to 10%)	2.2319	0.7047	2.54	0.011	1.2020	4.1442
J (10% to 20%)	1.7348	0.7227	1.32	0.186	0.7667	3.9253
K (20% to 30%)	0.6361	0.6768	-0.43	0.671	0.0790	5.1195
S (0% to -10%)	1.5065	0.4784	1.29	0.197	0.8084	2.8074
T (-10% to -20%)	1.1435	0.3721	0.41	0.680	0.6043	2.1639
U (-20% to -30%)	0.6361	0.2499	-1.15	0.250	0.2945	1.3739
Constant	0.1572	0.0281	-10.32	0.000	0.1106	0.2233

Table 7: Logistic regression *Sell* Returns (*I*, *J*, *K*, *S*, *T* and *U*)

The log likelihood in table 7 is reported as -295.4592. The number of observations is 693 and the Chi2 p-value is 0.0578. The positive return brackets are slightly easier to interpret in table 7. One can observe that the odds ratios are higher for positive returns. As commented under table 3 including both *Gain* and positive return brackets is not optimal. While most of the odds ratios are not significant, the interpretation would be the same. The return bracket *I*, for positive returns between 0 and 10%, is significant at the 5% level with a p-value of 0.011. Positive return brackets have higher odds ratios, indication that gains make the investor more likely to sell the stock. This again is in line with the previous results in the analysis. The odds ratios are illustrated in figure 10.

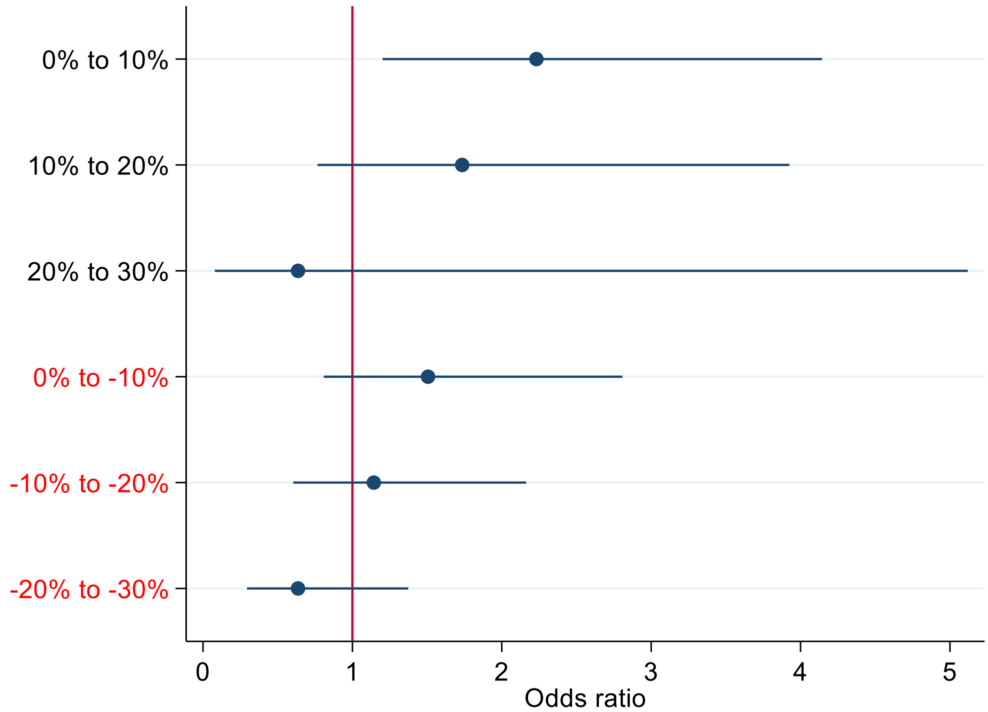

Figure 10: Odds ratios with corresponding 95% confidence interval 2

As one can see from figure 10, positive return brackets seem to have higher odds ratios than negative return brackets. Return bracket 0% to 10% are the only bracket that is significant on the 5% level. The positive and negative return brackets for 20% to 30% have the same observed odds ratio, but the 95% confidence level interval is much wider, this is because there are only 11 observations of stocks with positive returns between 20% and 30%.

5.2.6 Marginal effects

To better interpret the time element in the dataset, a marginal effect analysis is executed. Table 8 displays marginal effect analysis of the variable *Holdingtime*. The logistic regression model in table 5 is the base of the marginal effect analysis.

					95% Conf. Interval	
Margin at	Margin	SE	Z	P > I Z I	LL	UL
1 (Day 10)	0.1656	0.0338	4.90	0.000	0.0993	0.2318
2 (Day 60)	0.1628	0.0247	6.58	0.000	0.1143	0.2113
3 (Day 110)	0.1601	0.0172	9.26	0.000	0.1262	0.1940
4 (Day 160)	0.1574	0.0136	11.50	0.000	0.1306	0.1843
5 (Day 210)	0.1548	0.0162	9.55	0.000	0.1230	0.1866
6 (Day 260)	0.1522	0.0225	6.74	0.000	0.1079	0.1964
7 (day 310)	0.1496	0.0302	4.95	0.000	0.0903	0.2089
8 (day 360)	0.1471	0.0382	3.84	0.000	0.0712	0.2221

Table 8: Marginal effects

The marginal effects in table 8 is calculated from the *Holdingtime*. Margins 1-8 indicate the holding time starting at 10 days and ending at 360 days. The margins have 50-day intervals. Every marginal effect is highly significant, with p-values of 0.000. At 1 which is equal to day 10 in the dataset the probability of *Sell* = 1 is 0.1656. The chance of a 10-day old stock being sold is 16.56%. As one can see from table 8, there is a slight decrease in the marginal effect for each 50-day interval. For the last observation named 8 which is equal to day 360, a stocks odds of being sold is 0.1471 or 14.71%. One could believe that the marginal effect of *Holdingtime* should increase over time as more and more of the stocks included in the dataset are being sold. This would be the case, but stocks sold early (low value *Holdingtime*) are excluded from the dataset from the time they were sold. Therefore, a lower marginal effect on *Holdingtime* as days increases would indicate that the longer an investor holds a stock, the lower the chance of that stock being sold.

5.2.7 Likelihood test – Logistic model

Several likelihood tests are performed. The nested model is table 3 (m1). Likelihood test will be performed between all models. The results are displayed in table 9, 10, 11 and 12.

Assumption: m1 nested within m2	
LR Chi2(6)	6.89
Prob > Chi2	0.3313

Table 9: Likelihood test 1

Assumption: m1 nested within m3	
LR Chi2(7)	6.96
Prob > Chi2	0.4325

Table 10: Likelihood test 2

Assumption: m1 nested within m4	
LR Chi2(29)	25.24
Prob > Chi2	0.6660

Table 11: Likelihood test 3

Assumption: m1 nested within m5	
LR Chi2(5)	3.09
Prob > Chi2	0.6853

Table 12: Likelihood test 4

The log likelihood test is preformed between the models in table 3 and 4 (m2), the models in table 3 and 5(m3), the models in table 3 and 6 (m4) and the models in table 3 and 7 (m5). The p-values are 0.3313, 0.4325, 0.6660 and 0.6853 respectively. The high p-values indicates the data is consistent with the claim that the extra variables together do not substantially improve the model. An important note is that extra variables are tested together, not just individually (UCLA, 2021).

5.3 Cox proportional hazard model

The next part of the analysis of the disposition effect a cox proportional hazard model is utilized. The cox model will be advantageous when implementing the time variable into the data. Test and model are executed in STATA. Results are collected directly from STATA.

5.3.1 Hazard model – *Gain*

To begin the hazard model the data is declared as survival-time data. In table 13 the binary failure event is set as *Sell*, the time variable is set as *Holdingtime* and *Gain* is set as the independent variable.

Cox regression with Breslow method for ties						
No. of subjects	693				Observations	693
No. of failures	109				LR Chi2(1)	16.37
Time at risk	116 282				Prob > Chi2	0.0001
Log Likelihood	-603.8026					
					95% Conf. Interval	
Variable	Hazard Ratio	SE	Z	P > I Z I	LL	UL
Gain	2.3856	0.4873	4.26	0.000	1.5985	3.5604

Table 13: Hazard model - *Gain*

The total number of observations 693. In total there are observed 109 failures. Failures in this sense would be *Sell* = 1. The total time at risk is 116 282 days. The interpretation of this would be that the overall number of days stocks were hold is 116 282. The overall model is significant at the 1% level with a Chi2 p-value of 0.0001. The independent variable *Gain* is also significant at the 1% level with a p-value of 0.000. The log likelihood is -603.8026. The *Gain* hazard ratio is 2.3856. For hazard ratios above 1 the *Gain* = 1 group experiences a higher likelihood of the failure event, than those in group *Gain* = 0. In the case of this model the failure event is the stock being sold. If the stock is in state *Gain* = 1 the investor is 2.3856 times as likely to sell. When utilizing cox regression, the predicted effect of *Gain* = 1 is higher than that of the logistic regression. The high hazard ratio for *Gain* is further confirmation of the presence of the disposition effect.

5.3.2 Hazard model – Investor-fixed effects

In this model a dummy for each investor is added. The *Gain* variable is still included. The results are displayed in table 14.

Cox regression with Breslow method for ties						
No. of subjects	693				Observations	693
No. of failures	109				LR Chi2(1)	67.54
Time at risk	116 282				Prob > Chi2	0.0001
Log Likelihood	-578.2181					
					95% Conf. Interval	
Variable	Hazard ratio	SE	Z	P > I Z I	LL	UL
Gain	3.0036	0.7390	4.47	0.000	1.8544	4.8649

Table 14: Hazard model, Investor-fixed effects (Complete table displayed in appendix, table 22)

The model in table 14 is overall significant at the 1% level with a Chi2 p-value of 0.0001. The independent variable *Gain* is significant at the 1% level with p-value of 0.000. A few of the investor dummy variables are significant at the 5% level. Investors 5, 6, 24 and 26 have p-values of 0.013, 0.001, 0.041 and 0.033 respectively. These values are collected and displayed in the appendix. The *Gain* hazard ratio has further increased from the model in table 9. *Gain* now represents the effect of *Gain* on *Sell* using only within-investor variation. Hazard ratio of 3.0036 would indicate that the overall hazard of selling is 3.0036 times a likely when *Gain* = 1.

5.3.3 Likelihood test – Hazard model

A likelihood test is preformed. The model in table 13 and 14 are tested. The table 13 model is used as the nested model. The results are displayed in table 15.

Assumption: m1 nested within m2	
LR Chi2(29)	51.17
Prob > Chi2	0.0067

Table 15: Likelihood test 5

The Chi2 p-value is 0.0067. The results show that adding the *ID* dummy variables as predicators together results in a statistically significant improvement in the model fit.

5.3.4 Proportional hazards assumption test

Test of the proportional hazard assumption is completed for models displayed in table 13 and 14. These are displayed by table 16 and 17 respectively.

Proportional hazard assumption	
Time function: Analysis test	
Global test	
Chi2	0.50
df	1
Prob > Chi2	0.4795

Table 16: Proportional hazard assumption test 1

Proportional hazard assumption	
Time function: Analysis test	
Global test	
Chi2	20.11
df	30
Prob > Chi2	0.9138

Table 17: Proportional hazard assumption test 2

As indicated by table 16 and 17 the Chi2 p-values are 0.4795 and 0.9138 respectively. As mentioned, one of the key assumptions of the cox proportional hazard model is that the hazards are proportional. The p-value in table 12 and 13 indicate that the H0 hypothesis of proportional hazards cannot be rejected. The data indicates the hazards of the two groups are proportional (STHDA, 2023).

5.3.5 Graphical presentation – hazard model

To continue the analysis figures 11 and 12 are included. These tables display the survival rate over time and a Kaplan-Meier survival estimate.

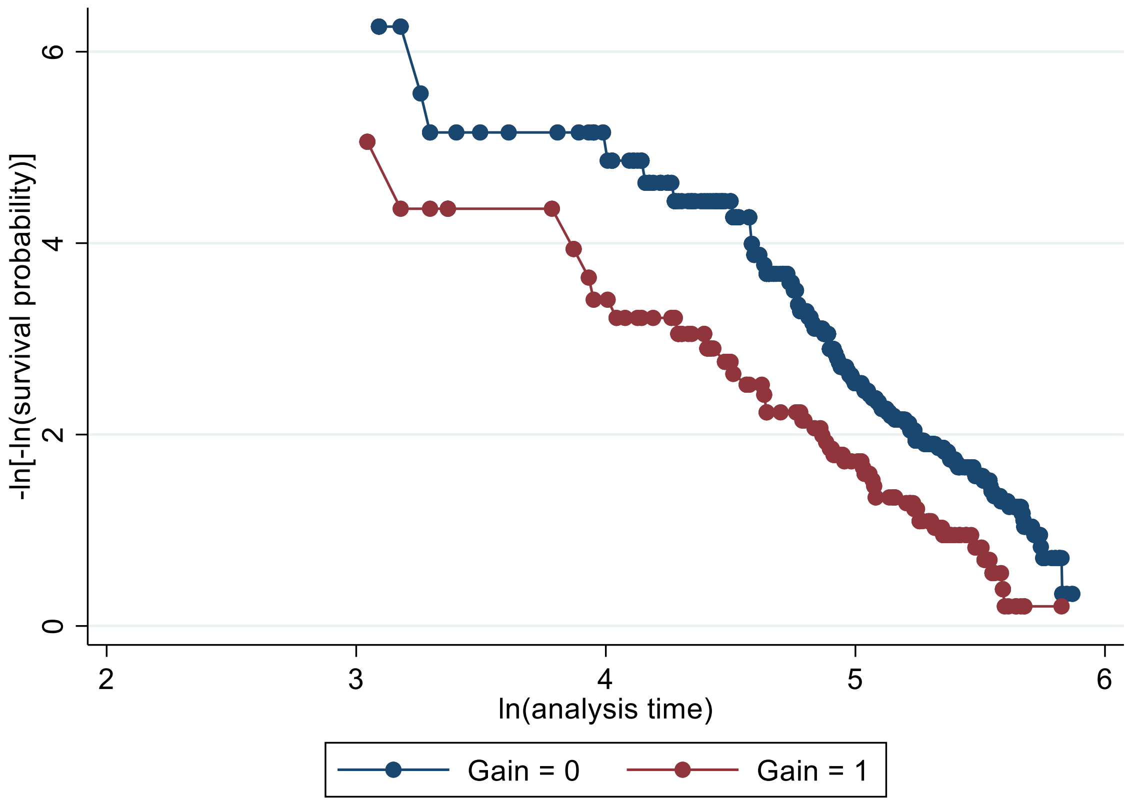

Figure 11: Cox survival probability

In figure 11 the survival probability over time is displayed in the log-log-plot. This graph further confirms the proportional assumption. As indicated by the graph the two lines are proportional. The interpretation of the graph is that the survival probability is lower if *Gain* = 1. Survival in this context is the failure event does not happen, ergo the stock is not sold. If *Gain* = 0 the survival probability goes up, indicating that the stock is more likely to survive, in other words being held.

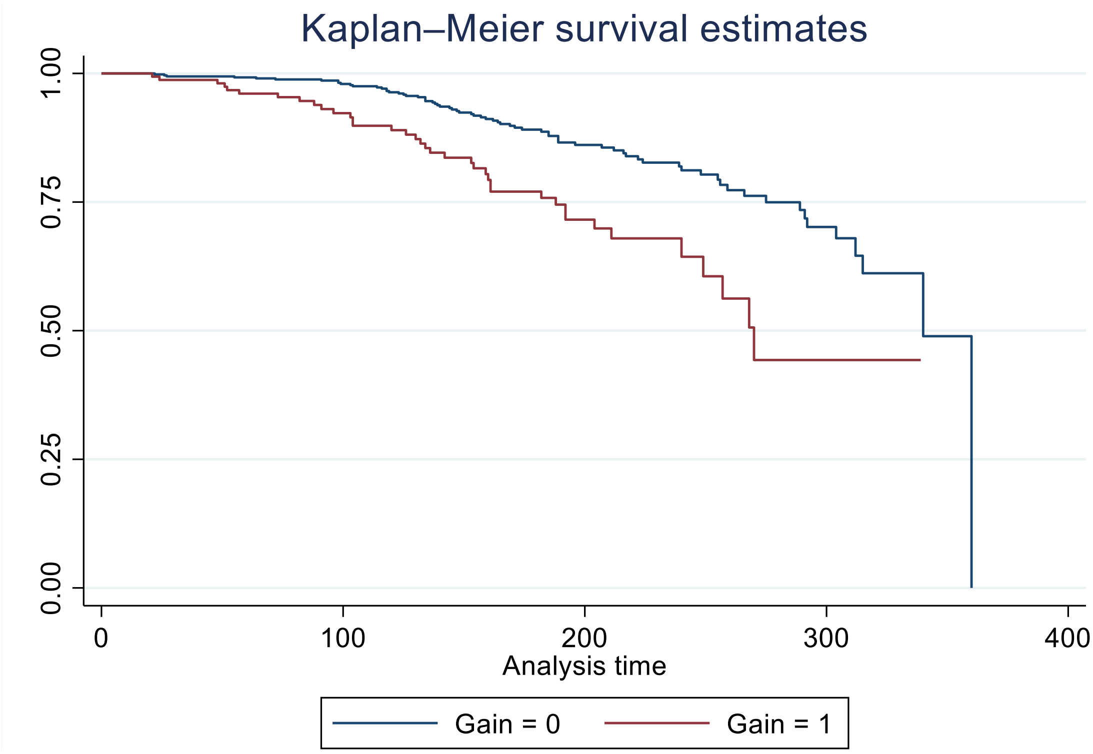

Figure 12: Kaplan-Meier survival estimate

In figure 12 a Kaplan-Meier survival estimate is displayed. The line illustrates percentage of stock that are sold over the given analysis time for groups *Gain* = 0 and *Gain* = 1. The sudden drop-of at days 360 is the end of the dataset with regards to *Holdingtime*.

5.4 Weibull regression

As mentioned in methodology, there is no constant element or probability of holding time when testing for the disposition effect. The final part of the analysis will include a survival time analysis using the Weibull distribution. The results of the Weibull survival time regression are displayed in table 18.

Cox regression with Breslow method for ties						
No. of subjects	693				Observations	693
No. of failures	109				LR Chi2(1)	15.61
Time at risk	116 282				Prob > Chi2	0.0001
Log Likelihood	-284.1205					
					95% Conf. Interval	
Variable	Odds Ratio	SE	Z	P > I Z I	LL	UL
Gain	2.3166	0.4675	4.16	0.000	1.5598	3.4406
Constant	~0	~0	-14.51	0.000	~0	~0
Ln p	0.8304	0.0774	10.72	0.000	0.6785	0.9822
p	2.2942	0.1777			1.9710	2.6704
1/p	0.4358	0.0337			0.3744	0.5073

Table 18: Weibull regression

The parameters of the model displayed in table 18 are the same as the previous survival time models. Number of subject and number of observations remain at 693, number of failures are 109 and the total time at risk is 116 282. The overall significance of the model is high. The model is significant at the 1% level with a Chi2 p-value of 0.0001. The *Gain* hazard ratio is measured at 2.3166 and it is also significant at the 1% level with a corresponding p-value of 0.000. The interpretation of the hazard ratio is equal to the previous survival time regression. An investor is 2.3166 times as likely to realize a stock if it is in state *Gain* = 1.

5.4.1 Akaike's and Bayesian information criterion

To test if the cox proportional hazard model or the Weibull regression has a better fit to the data the Akaike's and Bayesian information criteria is tested. The cox proportional hazard model in table 13 and the Weibull regression in table 18 are tested. The results of the tests are displayed in table 19 and 20.

Akaike's and Bayesian information criterion	
N	693
LL (null)	-611.9886
LL (model)	-603.8026
df	1
AIC	1209.605
BIC	1214.146

Table 19: AIC and BIC 1

Akaike's and Bayesian information criterion	
N	693
LL (null)	-291.9273
LL (model)	-284.1206
df	3
AIC	574.2411
BIC	587.8642

Table 20: AIC and BIC 1

The Weibull regression is tested in table 20. It produces far lower values for AIC and BIC. From this one can conclude that the Weibull regression model fits the data better than the cox proportional hazard model.

5.4.2 Graphical presentation – Weibull distribution

To better illustrate the survival rates of *Gain* = 0 and *Gain* = 1 the results are displayed graphically. The results are displayed in figure 13.

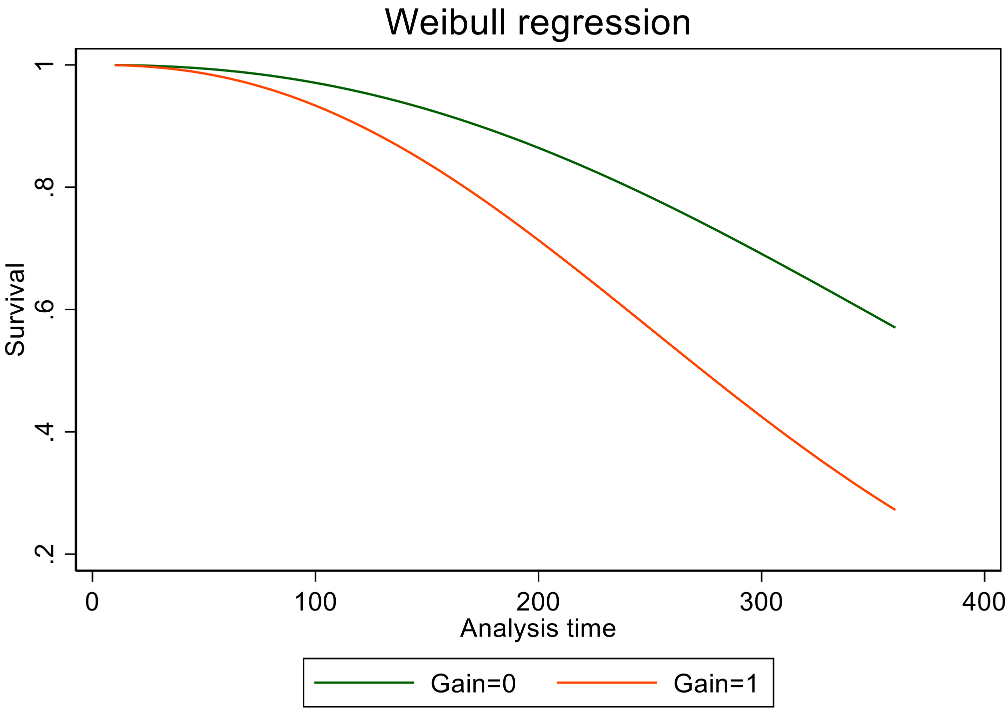

Figure 13: Weibull survival rate

As illustrated in figure 11. Stock with *Gain* = 0 has a far higher survival rate than stock with *Gain* = 1. It would also seem like there is a greater difference in survival rate as analysis time increases, but the overall trend in figure 11 is similar to the results displayed in figure 9. The overall survival of the stock is lower if *Gain* = 1.

6 Summary and discussion

In this chapter, I will gather the results from the last chapter and discuss these. Furthermore, I will point out certain weaknesses and challenges in the analysis. Lastly, I will point out possibilities for additional research.

6.1 Discussion of results

The analysis is primarily divided into three categories. The PGR and PLR, logistic regression, and cox proportional hazard model.

With regards to the PGR and PLR analysis, the results are in line with previous research on the subject. The disposition effect was measured at 0.103. The t-statistic was 2.804 with a corresponding p-value of 0.0089. The result is highly significant and indicates the presence of the disposition amongst the investors included in the dataset. The monthly values, excluding December, of the disposition effect are also positive. The corresponding significance of the values are rather low. Therefore, these results are not analysed further.

Logistic regression is utilized to further analyse the disposition effect. From table 3 the *Gain* odds ratio is measured at 2.0059 with a corresponding p-value of 0.002. Indicating that the investors in the dataset are over two times as likely to sell if the stock is in state *Gain*. This is in line with the general theory of the disposition effect and the results gathered in the PGR ang PLR part of the analysis. In the logistic regression more variables were added. In table 4 the first 3 return brackets of negative and positive returns are added. The *Gain* odds ratio remain significant, but the positive and negative return brackets are not significant. This was later accounted for in table 7, by excluding the *Gain* variable This led to more significant return bracket variables. In table 6 each investor is added as a dummy, this is equivalent to adding investor-fixed effects. The odds ratio for *Gain* in table 6 is estimated at 1.8754.

An important element in this research paper is the time element. In the cox proportional hazard model, the time element is more efficiently included in the analysis. In total 116 282 days are included in the hazard model. The hazard ratio of *Gain* is estimated at 2.3856. When adding investor dummies, the hazard ratio is estimated using only within-investor variation. This is illustrated in table 14, and the estimated hazard ratio is estimated at 3.0036. These results further confirm the disposition effect. The results are illustrated in figures 11 and 12. One can see a clear difference of the survival rates between groups *Gain* = 0 and *Gain* = 1.

Concluding the analysis part, a Weibull distribution is utilized. Table 18 reports a *Gain* hazard ratio of 2.3166. This result is in line with previous results in the analysis. Finally, the Weibull survival rate is displayed graphically in figure 11. Figure 11 indicates that there is a non-constant change in survival rate between the two groups *Gain* = 0 and *Gain* = 1.

6.2 Weaknesses

On to the subjects of weaknesses. One of the glaring weaknesses of this paper is the small number of observations. In total there are 244 purchasing decisions and 109 selling decisions across the 30 different investors. In terms of RG, RL, PG and PL there are 693 observations. Total number of days observed in the dataset are 116 282. The main findings in this paper are significant. When dividing the dataset into return brackets or analysing monthly, the results are simply too weak to conclude anything. Previous studies on the topic have had observations into the 100 million range (Pelster & Hofmann, 2018). The possibilities with observations in the 100 million range is far greater than the possibilities with the number of observations in this research paper. A good example of this in this paper would be the overall weak results when looking at the disposition effect month by month. Another one would be the return brackets variables. These variables contain few observations and are therefore difficult to include in the analysis.

An interesting point both under weakness and further research, would be the overall market conditions. With my analysis the overall trend of the market is not considered. The impact of positive or negative periods in the financial market is not considered, this could be a weakness and would as well be interesting for further research.

Another weakness is the fact that the data is collected by hand. There could be errors in reported values as well as some bias when it comes to me manually collecting the data.

6.3 Further research

My result is in line with previous studies on the disposition effect. The disposition effect is one of the more documented phenomena within behavioural finance. With a larger dataset there are several paths further research can take. As mentioned, one weakness in my paper is the small dataset. This led to insignificant results when looking at the data with a monthly perspective. With a larger dataset there is possibilities of grouping the investor into different categories (Pelster & Hofmann, 2018). Another possibility is further research on the monthly values of disposition effect. A previous paper by Terrance Odean on the subject has pointed out that investors are more willing to sell in December due to tax benefit (Odean, 1998).

Through Shareville, there is a function that allows for sorting investors across the Scandinavian countries. The countries included are Sweden, Denmark, Norway, and Finland. There are some studies on the disposition effect in Finland (Grinblatt & Keloharju, 2001). It would be interesting to compare the disposition effect across the Scandinavian countries. The data is available in Shareville.

Another part that would be interesting for further research is the return variables. In my paper there are in total 20 return variables. Because of the rather small number of observations these variables often resulted in insignificant results. With more data the analysis of the specific return bracket variables and analysis on the monthly basis would be far more significant.

7 Conclusion

This research paper is an empirical study of investors on the social trading platform Shareville in the period 01.01.2022 to 31.12.2022. The papers' main purpose is to detect the presence of the disposition effect amongst the investors. Furthermore, analyse predicators of the disposition effect. The study finds than positive and negative returns have a significant effect on investors selling decision. Positive returns make the investor far more likely to sell the stock.

In the PGR and PLR analysis I find that investors realize a larger proportion of gains than losses. This difference is highly significant, proving the presence of the disposition effect.

In the logistic regression the predicated odds show that investors are far more likely to sell the stock if it is in a state of positive return. The same result is also presented in the Cox proportional hazard model. Again, providing evidence that the investor is far more likely to realize the stock if it is in a state of positive return.

There is not enough data to say anything significant about the disposition effect on the monthly basis. There is observed a negative disposition effect in December, however there is not nearly enough data to conclude anything from that observation. Furthermore, the return bracket variables gave insignificant results.

8 Literature list

Fernano Boix. (2021). Er konklusionen pålitelig? Tidsskriftet. URL: https://tidsskriftet.no/2021/03/kommentar/er-konklusionen-palitelig

Jason Brownlee. (2020). Probabilistic Model Selection with AIC, BIC, and MDL. Machinelearningmastery. URL: https://machinelearningmastery.com/probabilistic-model-selection-measures/

David Cox. (1972). Regression Models and Life-Tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, *34*(2), 187–220. http://www.jstor.org/stable/2985181

Stephanie Glen. (2023). Likelihood-Ratio Tests (Probability and Mathematical Statistics). tatisticsHowTo. URL: https://www.statisticshowto.com/likelihood-ratio-tests/

Mark Grinblatt and Matti Keloharju. (2001). What Makes Investors Trade? *The Journal of Finance*, 56(2), 589–616. https://www.jstor.org/stable/222575

Adam Hayes. (2023). T-Test: What It Is With Multiple Formulas and When To Use Them. Investopedia. URL: https://www.investopedia.com/terms/t/t-test.asp

Rawley Heimer. (2016). Peer Pressure: Social Interaction and the Disposition Effect. *The Review of Financial Studies*, 29(1), 3177–3209. http://www.jstor.org/stable/44014952

David Hirshleifer. (2015). Behavioral Finance. *Annual Review of Financial Economics*, 7(1), 133-159. http://dx.doi.org/10.2139/ssrn.2480892

Daniel Kahneman and Amos Tversky. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, *47*(2), 263-291. https://www.jstor.org/stable/1914185

Jeffrey Kubik, Harrison Hong, and Jeremy Stein. (2004). Social Interaction and Stock-Market Participation. *The Journal of Finance*, 59(1), 137-163. http://www.jstor.org/stable/3694892

Nordnet. (2013). Nordnet köper sociala investeringsnätverket Shareville. URL: https://news.cision.com/se/nordnet/r/nordnet-koper-sociala-investeringsnatverket-shareville,c9430761

Nordnet. (2014). Investeringsnettverket Shareville med over 30.000 aksjeporteføljer. URL: https://news.cision.com/no/nordnet/r/investeringsnettverket-shareville-med-over-30-000-aksjeportefoljer,c9696846

Nordnet. (2023). This is Nordnet. URL: https://nordnetab.com/about/nordnet-overview/

Terrance Odean. (1998). Are Investors Reluctant to Realize Their Losses? *The Journal of Finance,* 53(5), 1775-1798. https://doi.org/10.1111/0022-1082.00072

Matthias Pelster and Annette Hofmann. (2018). About the fear of reputational loss: Social trading and the disposition effect. *Journal of Banking & Finance,* 94(C), 75-88. https://doi.org/10.1016/j.jbankfin.2018.07.003

Maarten van Rooij, Annamaria Lusardi, and Rob Alessie. (2011). Financial literacy and stock market participation. *Journal of Financial Economics,* 101(2), 449-472. https://doi.org/10.1016/j.jfineco.2011.03.006

Mark Seasholes, Lei Feng. (2005). Do Investor Sophistication and Trading Experience Eliminate Behavioral Biases in Financial Markets? Review of finance, 9(3), 305-351. http://dx.doi.org/10.2139/ssrn.694769

Shareville. (2023). Medlemmer. Collected 01.05.2023. URL: https://www.shareville.no/medlemmer

Hersh Shefrin and Meir Statman. (1985). The Disposition to Sell Winners Too Early and Ride Losers Too Long: Theory and Evidence. *The Journal of Finance, 40*(3), 777-790. https://www.jstor.org/stable/2327802

STHDA. (2023). Cox Model Assumptions. Statistical tools for high-throughput analysis URL: http://www.sthda.com/english/wiki/cox-model-assumptions

Magne Thoresen. (2017). Logistisk regresjon – anvendt og anvendelig. Tidsskriftet. URL: https://tidsskriftet.no/2017/10/medisin-og-tall/logistisk-regresjon-anvendt-og-anvendelig

UCLA. (2021). HOW CAN I PERFORM THE LIKELIHOOD RATIO AND WALD TEST IN STATA? Statistical Methods and Data Analytics. URL: https://stats.oarc.ucla.edu/stata/faq/how-can-i-perform-the-likelihood-ratio-wald-and-lagrange-multiplier-score-test-in-stata/

Figure 1. Liisa Uusitalo. (2009). Advertising models- an overview of effects on consumers. URL: https://www.researchgate.net/figure/asymmetric-value-function-of-prospect-theory_fig5_275028283

Figure 2. Llewelyn Fernandes. (2021). Understand the Logistic Regression Algorithm. URL: https://openclassrooms.com/en/courses/6389626-train-a-supervised-machine-learning-model/6405876-understand-the-logistic-regression-algorithm

Appendix, t-table. Shaun Turney. (2022). Student's t Table | Guide & Examples. URL: https://www.scribbr.com/statistics/students-t-table/

9 Appendix

Logistic regression					Observations	693
					Chi2	34.33
					Prob > Chi2	0.2678
Log Likelihood	-284.3891				Pseudo R2	0.0569
					95% Conf. Interval	
Variable	Odds Ratio	SE	Z	P > I Z I	LL	UL
Gain	1.8754	0.4710	2.50	0.012	1.1463	3.0683
ID						
2	0.6439	0.3830	-0.74	0.459	0.2006	2.0662
3	2.3377	2.1495	0.92	0.356	0.3855	14.1736
4	1.6277	0.9427	0.84	0.400	0.5230	5.0651
5	2.1852	1.4950	1.14	0.253	0.5716	8.3531
6	2.8292	1.9701	1.49	0.135	0.7226	11.0767
7	1.1099	0.6336	0.18	0.855	0.3625	3.3978
8	1.5520	0.8058	0.85	0.397	0.5609	4.2942
9	1.5191	1.3135	0.48	0.629	0.2789	8.2716
10	2.1852	2.6621	0.64	0.521	0.2007	23.7934
11	0.8275	0.4935	-0.32	0.751	0.2571	2.6631
12	1.0460	0.7504	0.06	0.950	0.2563	4.2680
13	2.1852	2.6621	0.64	0.521	0.2007	23.7934
14	1.4756	1.0988	0.52	0.601	0.3428	6.3509
15	0.8255	0.9281	-0.17	0.865	0.0911	7.4772
16	0.7558	0.4493	-0.47	0.638	0.2357	2.4234
17	4.3152	3.5343	1.79	0.074	0.8666	21.4870
18	1.8269	1.3665	0.81	0.420	0.4217	7.9143
19	0.6393	0.4083	-0.70	0.484	0.1828	2.2352
20	1.4201	1.2107	0.41	0.681	0.2670	7.5513
21	2.6035	3.1558	0.79	0.430	0.2419	28.0116
22	1.6221	1.4019	0.56	0.576	0.2981	8.8260
23	0.7005	0.4163	-0.60	0.549	0.2185	2.2455
24	0.9786	0.8383	-0.03	0.980	0.1825	5.2460
25	3.4897	3.2936	1.32	0.185	0.5488	22.1905
26	0.7824	0.5548	-0.35	0.729	0.1949	3.1408
27	0.9981	1.1613	-0.00	0.999	0.1020	9.7620
28	1.7662	1.1935	0.84	0.400	0.4697	6.6415
29	4.4972	3.1152	2.17	0.030	1.1569	17.4816
30	3.1239	2.0622	1.73	0.084	0.8566	11.3924
Constant	0.1280	0.0472	-5.57	0.000	0.0621	0.2637

Table 21: Logistic regression – Investor-fixed effects (complete)

Cox regression with Breslow method for ties						
No. of subjects	693				Observations	693
No. of failures	109				LR Chi2(1)	67.54
Time at risk	116 282				Prob > Chi2	0.0001
Log Likelihood	-578.2181					
					95% Conf. Interval	
Variable	Odds Ratio	SE	Z	P > I Z I	LL	UL
Gain	3.0036	0.7390	4.47	0.000	1.8544	4.8649
ID						
2	0.4840	0.2807	-1.25	0.211	0.1553	1.5085
3	0.5832	0.4633	-0.68	0.497	0.1229	2.7674
4	0.7391	0.3997	-0.56	0.576	0.2560	2.1333
5	4.5527	2.7651	2.50	0.013	1.3844	14.9711
6	7.4613	4.6380	3.23	0.001	2.2065	25.2306
7	1.1029	0.5858	0.18	0.854	0.3894	3.1237
8	1.3266	0.6275	0.60	0.550	0.5249	3.3525
9	0.6934	0.5431	-0.47	0.640	0.1493	3.2191
10	0.3684	0.3954	-0.93	0.352	0.0449	3.0199
11	0.4060	0.2282	-1.60	0.109	0.1349	1.2217
12	0.3703	0.2500	-1.47	0.141	0.0985	1.3912
13	2.8580	3.0396	0.99	0.323	0.3554	22.9806
14	0.4828	0.3340	-1.05	0.293	0.1244	1.8735
15	0.1788	0.1902	-1.62	0.106	0.0222	1.4391
16	0.4638	0.2627	-1.36	0.175	0.1527	1.4079
17	1.9213	1.2850	0.98	0.329	0.5179	7.1267
18	0.5989	0.4056	-0.76	0.449	0.1588	2.2587
19	0.3983	0.2445	-1.50	0.134	0.1196	1.3265
20	1.1236	0.8813	0.15	0.882	0.2415	5.2275
21	0.4695	0.4989	-0.71	0.477	0.0585	3.7676
22	0.8703	0.6819	-0.18	0.859	0.1873	4.0425
23	0.7016	0.3930	-0.63	0.527	0.2340	2.1035
24	0.1941	0.1556	-2.04	0.041	0.0403	0.9342
25	1.4503	1.1415	0.47	0.637	0.3101	6.7831
26	0.2353	0.1595	-2.13	0.033	0.0623	0.8887
27	0.4013	0.4276	-0.86	0.392	0.0497	3.2403
28	2.5385	1.5498	1.53	0.127	0.7671	8.3996
29	0.6891	0.4343	-0.59	0.555	0.2003	2.3704
30	1.3177	0.7454	0.49	0.626	0.4348	3.9936

Table 22: Hazard model – Investor-fixed effects (complete)

cum. prob	$t_{.50}$	$t_{.75}$	$t_{.80}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$	$t_{.999}$	$t_{.9995}$
one-tail	**0.50**	**0.25**	**0.20**	**0.15**	**0.10**	**0.05**	**0.025**	**0.01**	**0.005**	**0.001**	**0.0005**
two-tails	**1.00**	**0.50**	**0.40**	**0.30**	**0.20**	**0.10**	**0.05**	**0.02**	**0.01**	**0.002**	**0.001**
df											
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.000	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.000	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.000	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.000	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.000	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.000	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.000	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.000	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.000	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.000	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.000	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.000	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.000	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.000	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.000	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.000	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.000	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.000	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.000	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.000	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.000	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.000	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.000	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.000	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.000	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	0.000	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	0.000	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232	3.460
80	0.000	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195	3.416
100	0.000	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174	3.390
1000	0.000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098	3.300
z	0.000	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.090	3.291
	0%	50%	60%	70%	80%	90%	95%	98%	99%	99.8%	99.9%
	Confidence Level										

Table 23: T-value table

Investor	Name of the Investor and name of portfolio
ID	Value 1 to 30 given individually to every investor
Date	Date of selling decision.
Stock	Name of stock
Holdingtime	Holding time in days since purchase
Sell	1 if stock is sold, 0 if stock is not sold
Return	Return of stock
Gain	1 if stock return is positive, 0 if stock return is negative.
I (0% to 10%)	1 if return is between 0% and 10%, 0 otherwise.
J (10% to 20%)	1 if return is between 10% and 20%, 0 otherwise.
K (20% to 30%)	1 if return is between 20% and 30%, 0 otherwise.
L (30% to 40%)	1 if return is between 30% and 40%, 0 otherwise.
M (40% to 50%)	1 if return is between 40% and 50%, 0 otherwise.
N (50% to 60%)	1 if return is between 50% and 60%, 0 otherwise.
O (60% to 70%)	1 if return is between 60% and 70%, 0 otherwise.
P (70% to 80%)	1 if return is between 70% and 80%, 0 otherwise.
Q (80% to 90%)	1 if return is between 80% and 90%, 0 otherwise.
R (+90%)	1 if return is over 90%, 0 otherwise.
S (-0% to -10%)	1 if return is between -0% and -10%, 0 otherwise.
T (-10% to -20%)	1 if return is between -10% and -20%, 0 otherwise.
U (-20% to -30%)	1 if return is between -20% and -30%, 0 otherwise.
V (-30% to -40%)	1 if return is between -30% and -40%, 0 otherwise.
W (-40% to -50%)	1 if return is between -40% and -50%, 0 otherwise.
X (-50% to -60%)	1 if return is between -50% and -60%, 0 otherwise.
Y (-60% to -70%)	1 if return is between -60% and -70%, 0 otherwise.
Z (-70% to -80%)	1 if return is between -70% and -80%, 0 otherwise.
AA (-80% to -90%)	1 if return is between -80% and -90%, 0 otherwise.
AB (-90% to -100%)	1 if return is between -90% and -100%, 0 otherwise.

Table 24: Variable list and definitions

NTNU

Norwegian University of
Science and Technology