Aalto University
School of Business | Mikkeli Campus

EMPIRICAL ANALYSIS OF EFFECTS OF VIX ON RETAIL INVESTOR ACTIVITY DURING THE COVID-19 PANDEMIC

Evidence from Robinhood users

Hugo Uusi-Mikkola

International Business
Bachelor's Thesis
Instructor: Roman Stepanov
Date of approval: 9th April 2021

Aalto University
School of Business
Bachelor's Program in International Business
Mikkeli Campus

Aalto University
School of Business | Mikkeli Campus

EMPIRICAL ANALYSIS OF EFFECTS OF VIX ON RETAIL INVESTOR ACTIVITY DURING THE COVID-19 PANDEMIC

Evidence from Robinhood users

Hugo Uusi-Mikkola

International Business
Bachelor's Thesis
Instructor: Roman Stepanov
Date of approval: 9th April 2021

Aalto University
School of Business
Bachelor's Program in International Business
Mikkeli Campus

AALTO UNIVERSITY
SCHOOL OF BUSINESS
Mikkeli Campus

ABSTRACT OF
BACHELOR'S THESIS

Author: Hugo Uusi-Mikkola
Title of thesis:

Date: 9 April 2021

Degree: Bachelor of Science in Economics and Business Administration

SUPERVISOR: ROMAN STEPANOV

Objectives

The main objectives of this study were to deepen the understanding of retail investor activity under different market conditions, and to see if the VIX provides predicting power over retail investor activity. The different market conditions were separated as the full timeframe, before the pandemic, and during the pandemic. Secondary object was to shed light on how retail investors perceive VIX, as measure of fear or as a measure of noise.

Summary

This study collected 3-years' worth of data from VIX and popularity data from retail brokerage Robinhood to measure retail investor behavior between May of 2018 and August of 2020. The findings were analyzed based on various charts of the datasets and regression analysis divided into three timeframes based on level of VIX.

Conclusions

The conclusions of this study are that during market turmoil and higher level of VIX, variation in VIX provides some predictability over retail investor activity, whereas the relationship before the pandemic is non-existent. The results point to VIX being a proxy for noise, as during higher levels of VIX a sharp change to either direction resulted into an increase in retail investor activity with a delay. From the basis of the findings this study, VIX is a viable option as a variable to provide additional explanatory power over retail investor activity in a more comprehensive model.

Key words: Robinhood, Retail investors, VIX, Investor behavior, Covid-19, Stock market, Noise trading, Herding effect

Language: English

Grade:

COVER PAGE
TITLE PAGE
ABSTRACT
TABLE OF CONTENTS

1. INTRODUCTION

The retail investors have been seen as low-power force in the market for a long time, and with the development of technology, and investing platforms and continuously reduced barriers to investing have significantly increased the amount of retail investors in the market. The latest developments happening in the beginning of the novel coronavirus Covid-19 pandemic during the year 2020, when the markets experienced a surge of retail investors, extreme market movements and higher levels of VIX, than during the financial crisis of 2008. A bulk of literature has shown increased economic significance within the past decade (Barber and Odean, 2013; Boehmer, Jones and Zhang, 2016; Aharon and Quadan, 2020; Ballinari, 2020; Barber et. al, 2020; Bates, 2020 and others)

Retail investors are traditionally an irrational force in the market and the tracking of the behavior has been a difficult task. Retail investors have proven to show systematic patterns in their behavior, but a there is a void in literature around models for predicting retail investor behavior, due to the irrationality of investors and the significance of multitude of signals to their behavior. This thesis focuses its research to examine retail investor behavior from the online brokerage Robinhood's data and to see if VIX, the index of implied volatility of S&P500, could offer predicting capabilities of retail investor activity in the market. VIX has two purposes in this thesis, to separate the timeframe between extreme market conditions, and regular market conditions, and to analyze how changes in VIX affect retail investing activity during these different timeframes.

1.1 Background and Research Problem

Retail investor activity has grown with accelerating speed with the development of technology, smart devices and internet. The development has continuously reduced the barriers of entry to investing, the latest development being the removal of trading commissions. Growing number of retail investors has increased the economic value for assessing the behavior and activity of retail investors. During the year 2020 and the on-going novel coronavirus Covid-19 pandemic retail investors and peculiar market moves made it to headlines and the interest for forecasting the behavior has increased.

Robinhood crowd and the amount of retail investors has grown during the pandemic and the market has seen higher levels of fear than measured before (Ramelli and Wagner; 2020). Retail investor behavior has been a focus of multiple studies during the 2000's and the Pandemic combined with the increase of retail investors in the market offers a great opportunity to better understand the behavior in context of established market measures. So far, the relationship has been studied in context of various variables, including market returns, forecasting capabilities of different variables still lacks research, and this thesis seeks fill that void within the literature and to improve on the knowledge of tracking retail investor activity.

This study researches the usability of implied market volatility index VIX to predict retail investor activity. The timeframe being studied is in three parts, as a whole, before and during the pandemic, after which the results are contrasted between each other and then further with findings of previous literature.

1.2 Research Questions

The research questions to be first addressed in the literature review and further researched in this thesis are as follows:

1. Does VIX provide predicting power over retail investor activity?
2. How does the level of VIX affect retail investor activity?
3. How retail investors account for variation in VIX under different market conditions?

1.3 Research Objectives

While answering to these research questions, this thesis seeks to shed light on retail investor activity in different market conditions and research the relationship between VIX and retail investor activity. In addition, examine how the relationship changed during the covid 19-pandemic, compared to regular market conditions before 2020. And finally, to improve the knowledge for tracking retail investor activity and to deepen the understanding of retail investor behavior under different market circumstances.

2 LITERATURE REVIEW

The purpose of this literature review is to consult previous academic literature published from the topic of financial theory, more specifically the ideas of Efficient market hypothesis and the weak form of efficiency. From the topics of behavioral finance theory, the literature review introduces concepts such as noise trading, the Herding effect, as well as various empirical studies on behavioral finance, to deepen the understanding of the research topic. Empirical aspects to be reviewed include retail investor behavior as well as the effects of investor sentiment and volatility to the markets. Building upon these aspects, the goal is successfully conducting a quantitative research to analyze retail investor behavior in relation to VIX before, and during the covid-19 pandemic.

Firstly, to understand investor behavior the literature review addresses academic literature of financial theory and behavioral finance, and introduces crucial elements, such as EMH to understand the traditional ideas of efficient market. To further understand investor behavior and predictability of retail investors, aspects of behavioral finance, for example the Herding effect and noise trading, are introduced. Secondly, to illustrate the behavior in different circumstances the literature review assesses empirical studies around retail investor behavior and how technology has affected retail investors trading behavior. Retail investor activity is discussed simultaneously to provide evidence of the theories in the behavior and the significance of retail investors in the market.
Lastly the literature combines academic literature of market behavior during the covid-19 pandemic to further understand the effects on retail investor activity, investor sentiment and the market during the timeframe to be studied.

2.1 Efficient Market Hypothesis

Efficient Market Hypothesis (EMH), introduced by Fama (1970) is a classic financial theory, which relies on four conditions summarized as:

I. The public availability of information
II. The speed with which this information can be absorbed and lead to a new price equilibrium

III. Investor self-interest
IV. Investor rationality and the extent to which investors exhibit effective and efficient cognitive behavior

The basic assumption behind EMH is that markets are efficient relative to information available, unless this information offers an opportunity to the investors trading on its basis to earn abnormal returns. Thus, if information is universally available, investors are not able to consistently earn abnormal returns. This theory has been basis to financial theory for decades and has emerged criticism against it, which argues that for the markets to be efficient, all investors should be able to assess the information and make the right choices based on it, thus being rational in their trading decisions. The responses to this criticism are those of irrational investors are lacking systematic patterns, thus eventually canceling each other out. The idea for this is that of random walk, that with price movement there is a stochastic process with independent identically distributed binomial random variables, or that the stock prices move independently and randomly in the short run (Andrikopoulos, 2005). Second factor is arbitrage. Dybvig and Ross (1989) give the following example of what arbitrage is: "A simple example of Arbitrage is the opportunity to borrow and lend costlessly at two different fixed rates of interest." – thus arbitrage exists in imperfect market where price deviations occur, caused by irrational investors – or noise investors, but are eventually canceled by competing arbitrageurs. EMH is a basis from which behavioral finance has been built upon, to be able to research investor behavior in different market conditions, it is crucial to know how the market has been seen and what is the ideal financial environment (Andrikopoulos, 2005). Out of different forms of efficiency, in this thesis, the focus is on the weak form of market efficiency, or as Malkiel (1989) explains, it is the idea that historical prices cannot be devised to earn abnormal returns, which is closely related to the ideas of random walk within the market.

Financial theory tries to explain the markets rationally, whereas behavioral finance combines psychology and real-world events with investing and finance. This thesis focuses on explaining market behavior based on existing market measure VIX, thus the idea of weak form of efficiency – conducting research if VIX includes information of retail investor activity.

2.2 Behavioral Finance

Behavioral finance theory has built itself on the arguments against the EMH and the ideal market, as the world very rarely is at ideal state. The field of behavioral finance combines psychological theory to economic and financial theory and investing. Hirshleifer (2001) reviews theories around psychology and asset pricing theories and shows that the field lacks properly implementable theory or models. The behavioral finance relies more on the empirical findings of investor behavior under different circumstances and on different basic psychological theories applied to investing. In other words, behavioral finance is a combination of psychological theories and biases individual people have in their financial and economical behavior (Bates, 2020). The idea of this thesis is similar, to analuyse the results of an empirical study of retail investors being conducted, and to draw that to previous behavioral finance context to understand the relationship.

Individual investors are subject to many psychological biases and the behaviors negatively affect the returns of their investments. According to Barber and Odean (2013) one of the more prominent biases is disposition effect, that investors tend to hold on losing stocks while selling the winning stocks, which often increases the tax liability. Also, large pool of retail investors holds an under-diversified portfolio with unnecessary risk and with ignorance to costs related to their investments, which is contrarian to the ideas behind EMH of rational investor behavior and traditional pricing models of diversified portfolios and systematic risk. Investors in the real world tend to take part in momentum and popularity and act based on irrational noises, though as group, systematically (Schmittmann et al., 2014). These behaviors affect the market and give the possibility for informed investors to take advantage, as for every sale there is a buyer, for every loss there is a winner (Boehmer et.al 2020). These aspects of investor behavior are beneficial for explaining empirical results, and in the context of this study, these offer a pathway to explanations of observed findings.

2.2.1 Noise traders

Noise traders are irrational investors acting on noise, or different signals that catch their attention. The term “noise traders” has been used in literature previously interchangeably with retail investors, but this significantly undermines the magnitude of retail investors in the market. Foucault et. al (2011) define noise traders as retail

investors that act based on non-informational reasons. Irrational investors cause price deviation in the market, but also increase liquidity and trading volume by their actions, which leads to balancing force in the market. This is on par with the traditional idea of noise trading, that the increased liquidity helps the market to balance closer to fundamental value. However, a bulk of recent studies have illustrated more systematical behavior from retail investors (Economou, Philippas and Caporale, 2008; Barber and Odean, 2013; Aharon and Quadan, 2020; Ballinari, 2020; Barber et al., 2020).

There is clear evidence that retail investors are not behaving absolutely irrationally, but behave somewhat systematically in some conditions, but the signals behind the activity can be defined as noise in a sense that the short attention of an individual investor, steers them towards the decision that “keeps the most noise” and catches the attention. Hirshleifer (2001) discusses asset pricing theories based on imperfect rationality, pure noise trading combined with positive feedback trading cause overreaction and hence negative returns in the future, as the price corrects. Positive feedback trading is a phenomenon where investors form an idea of the trend of a stock and exaggerate it, leading to increased volatility and negative returns. This explains herding on a theoretical level, where the investors get over-confident and seek momentum based on positive price-reactions or news, and not fundamentals behind the price.

2.2.2 Herding Effect

Herding effect at its simplest is a group of investors collectively buying or selling a stock or a security, causing price deviation from the fundamental value. Continuing from the previous section, the theory of herding can be linked to positive feedback trading, where investors are over-confident of expected returns or base the trading decisions to extrapolating trends or previous success of a security (Hirshleifer, 2001). Traditional approach is that herding events happen unexpectedly and from irrational signals and it is a systemic pattern of retail investor behavior, which can be linked to momentum seeking behavior, where the investors seek to benefit from a certain momentum a stock, a security or a market has (Christie and Huang, 1995). Herding effect is often related to short attention span of retail investors, as well as the

irrationality of investors, where the investment decision is made upon popularity of a specific stock, rather than the fundamentals behind the price.

Christie and Huang (1995) find herding to happen especially during market stress and exaggerated price movements. Economou et. al (2008) further show herding to be more prone to positive market movement and is not limited to extreme market movements. Barber et. al (2020) found that herding effect causes abnormal prices and shortly after herding effect, negative returns as the increased price encourages people to sell, and the lower demand drops the price. Deducting from the basis of herding effect, increase of retail investors in the market would affect the volatility of the market. This thesis employs the idea that macro level systematic behavior, similar to herding effect of retail investors exists, and explaining the market conditions may offer a pathway to further understand how retail investors take part in the market.

2.3 Retail investor behavior

Retail investors have formed a force in the market that has been often defined being noise traders, irrational force that trades based on limited attention, and acts based on popularity and sentiment (Barber and Odean, 2008; Burghardt et.al, 2008; Foucault et. al 2011). Retail investors, have been always been an underdog in the market compared to institutional investors, only being able to cause small price anomalies due to noise trading, or otherwise irrational trading behavior.

Recent empirical studies (Boehmer et. al, 2016; Ekkayokkaya et. al, 2020; Moss et. al, 2020; Pagano et. al, 2020) have given contrarian results, showing retail investors being more significant force in the market, especially under more extreme market conditions, than they have been recognized previously and compared to institutional investors, retail investors pay attention to different information. Despite this, it is proven that retail investors are more prone to multitude of signals, such as weather and with limited attention of retail investors, the magnitude of signal affects the trading activity (Schmittmann et. al, 2014; Barber and Odean, 2008). The problem with the systematic behavior of retail investors is that the strategies are rarely purely related to fundamentals behind the price, but outside signals, in which case the increase of retail investors suggest more drastic price deviations. Change in technology and investing

platforms has changed the way we should address retail investors in the market, as the continuous increase is only accelerating, especially during the pandemic.

2.3.1 Online Trading

The development of technology has continuously removed barriers for investing. Online trading has made trading easier for retail investors, since the change from using brokers via phone calls and paying significant trading fees, the investors are able to make timely trades with fraction of the fees. Online brokerages have challenged classic brokerages by cutting costs and improving on the usability (Barber and Odean, 2001). Easier trading has raised concerns on possible increasing noise trading with the increase of retail investors. Online platforms offer timeliness and higher spectrum of investing opportunities to retail investors, but at the same time individual's capability to assess available information is not sufficient to choose always the best option, thus showing that according EMH these activities offer opportunities for arbitrageurs.

During recent years with the technology development, investing has become continuously easier for individuals, as all major brokerages offer a smartphone application for trading. And to even further reduce the barriers, brokerages have significantly decreased commissions, and some brokerages such as Robinhood even offer commission-free trading. The ease and democratization of trading has increased the number of retail investors in the market, which was even further accelerated during the covid-19 pandemic

2.3.2 Robinhood investors

Robinhood (RH) is an online brokerage, which was the first to offer commission-free trading, this combined with attractive and simplified user interface has attracted significant amount of beginner retail investors. RH has been able to gain significant user base, especially during 2020, when the user base surpassed 13 million in May 2020. The covid-19 outbreak and the following market crash attracted many beginner investors to take advantage of the "dip". Users value the simplicity and the lack of barriers to trading on RH application. (Barber et. al, 2020; Rooney 2020). Robinhood

has been related to disruptive investor behavior and there is evidence on obscure trading choices such as being ACB (Aurora Cannabis) being the most held stock on Robinhood surpassing companies such as AAPL (Apple) (Welch, 2020).

RH investors have been related to herding events, which are followed by negative returns in short-term. Herding behavior can be linked to the simplified user interface of RH showing limited data to the users (Barber et. al, 2020). Relating to the pure noise trading model discussed by Hirshleifer (2001) that herding may be seen as combination of noise trading and positive feedback trading, which is only further exploited by RH.

Retail investors on RH have proven to wield significant power in the market and that these investors are taking advantage of momentum as well as contrarian strategies or “buying the dip” (Pagano et. al 2020). Boehmer et. al (2016) find systematic patterns in retail investor behavior marketable retail order flows may predict future returns, which further gives proof of exploiting contrarian strategies. The contrarian strategies as well as the herding events can be directly linked to the limited attention of retail investors and that purchase decisions are made from small subset of stocks that catch the attention of the retail investor (Barber and Odean, 2007). Robinhood gives users simplified charts, such as “Top 20 Movers” which indicates any drastic moves in the market, grabbing the attention of the users. The exploitation of limited attention and increased userbase suggests increased noise investing in the market as well as surge in irrational movement, thus relating to increased fear and volatility levels.

2.3.3 Following retail investor behavior

Following retail investor behavior in the market is not a simple trick and there have been multiple methods for empirical studies. Boehmer et. al (2016) recognize retail trades from TAQ (New York Stock exchange Trade and Quote database) by differentiating retail order flows with them ending in fraction of penny amounts due to regulatory conditions. Nicolaus (2010) uses dataset extracted from German secondary market. Chiah and Zhong (2020) uses global data from Datastream to measure global changes in trading volume during the Covid-19 pandemic. Other often used way to recognize retail investor is to use data from a retail investor brokerage. Desagre and

D'Hondt (2020) used data from a Belgian brokerage and Barber and Odean (2001) used data from Ameritrade.

Recent surge of users on Robinhood as well as open access to popularity data (Robinhood closed the open access in August of 2020) has encouraged many studies to use data from Robinhood (Barber et. al, 2020; Moss et. al, 2020: Pagano et. al, 2020; Welch, 2020). This thesis will use the open access popularity data from Robinhood for the analysis for it being proxy for retail investor activity and due its increased popularity, it offers wide sample of retail investors.

2.4 Volatility of returns

Volatility of market returns, measures market uncertainty, risk and investor sentiment in the market. VIX is the best-known index of implied volatility of returns, the calculation is based on option trade of the underlying index SPX (S&P500). The same calculations are used in other indexes as well, for example the VDAX-NEW replicates the calculation of VIX in the German market. Volatility of returns traditionally is a measure of expected risk in the market, but also works as proxy of investor sentiment, which relates to the name "fear-gauge" as in when investors are expecting turmoil in the market, the VIX levels increase, as in VIX level inversely explains investor sentiment (Whaley, 2000).

2.4.1 VIX-index

VIX-index measures investors' implied expected volatility of market returns in the next 30 days of SPX. VIX index has shown to have multitude of indications of the market and investor behavior. VIX was originally introduced by Whaley (1993) and it is calculated by The Chicago Board Options Exchange (CBOE). Originally it was calculated based on S&P 100 options trade but was updated to use S&P 500 in 2003 based on its higher popularity. The index is calculated by the ratio of market call and put options (both at-the-money as well as out-of-the-money options) (Potter and Whaley, 2008). The mentioned "fear-gauge" -name comes from when investors begin expecting negative results, sales of index put options increase, showing more significant increases in VIX, compared to expected growth in the market. VIX is

popular in academic studies as well as in media coverage due to its public access on the CBOE website, for example Choi (2019), Sarı and Kartal (2020) as well as Ramelli and Wagner (2020) used VIX in their studies.

Lei et. al. (2012) use VIX as a proxy for investor sentiment due to its qualities to increase when investors are more fearful in the market, as well as in relation to behavioral finance, noise traders are often over-confident, higher VIX-levels translate to higher noise and VIX is correlated with trading volume. Higher volatility is also a proxy for market risk, and any relation with increase in investor activity during higher volatility levels suggest willingness higher risk taking. Aharon and Quadan (2020) researched correlation between VIX and retail investors attention to their trading platforms and discovered that greater changes in VIX correlate with investors' attention to their trading platforms. These examples would suggest usability for VIX as a proxy for noise in the market.

2.5 Stock Markets and the Covid-19 Pandemic

The global effect of the covid-19 pandemic in the past year is evident to everyone, and the reactions around the world have been significant, the stock market as a whole with individual highlights and moments. The combined power of RH crowd during the covid-19 pandemic has been a topic of news headlines and the power is evident, but the signals and rationality is still somewhat unclear. Peculiar events, which could be defined as extreme herding and attention driven events, during the pandemic include the surge of Hertz (HTZ) stock after filing for bankruptcy with massive retail investor attention and extreme surge of GameStop (GME) stock in early 2021. Both of these events were witnessed, and fueled, from an investing subreddit r/wallstreetbets, which especially during the latter, gained global media attention. The events of the year 2020 offer an opportunity to research how retail investors, as people who have had to spend more time home, have taken part in the stock market and how that is related to different factors and indices.

Albulescu (2020) was an early study on covid-19 pandemic spreading globally. The results of the spreading disease began showing in the markets and the death ratio outside of China correlated positively with VIX, implying increased fear in investors.

Sarı and Kartal (2020) further showed correlation with gold prices, often seen as implication of increased instability in the market, as gold is seen as low-volatility asset. Ramelli and Wagner (2020) and Talwar et. al (2021) provided further insight on the financial and economic effects of covid-19 and how it caused intense reactions to stock prices and high fluctuation in individual stocks, industries and even market wide. In March 2020 the VIX reached its all-time-high, surpassing the levels of financial crisis in 2008. Black swan -event describes well the pandemic, as there has not been such widespread outbreak causing as drastic movements in the market.

The interest to stock market among retail investors has risen during the pandemic and has been an affecting market force throughout the year 2020. Retail investors provided liquidity to a volatile market and took risks in their contrarian investment strategies (Ritual, 2020). The contrarian strategies are evidence of systematic behavior of retail investors in the market having a significant effect. However, high VIX-levels with translate to increased “fear” in the stock market and increased attention from retail investors suggests noise induced behavior and willingness for higher risk. Ballinari (2020) found that following retail investors sentiment provided improved forecasting capabilities during the covid-19 pandemic, relating it to continuously growing number of retail investors, and thus, increasing power they wield in the market. The rationality of this power is arguable. Contrarian and momentum seeking strategies can both be linked to limited attention of the investors, and brokerages being able to take advantage of that attention, may have unprecedented power in steering these forces.

2.6 Conclusions

The limited resources of individual investors relate to their behavior, but empirical evidence shows that retail investors have a growing power in the market and show systematic actions in the market providing economic significance. Further so, the increasing number of retail investors in the market during the pandemic offers a valuable opportunity to find patterns in the behavior.

Under extreme conditions the retail investors may provide liquidity from “buying the dip”, but also cause peculiar attention driven price deviating herding events to individual stocks and drive the volatility in seek of momentum. The value of understanding the rationale behind the behavior of increasing amount of retail

investors is becoming higher for arbitrageurs and institutional investors. This thesis seeks to shed light on the retail investor activity during the pandemic in relation with the high levels of fear and increased amount of retail investors in the market.

VIX index provides a proxy for market sentiment and during 2020 it reached highest levels in its measuring history, translating to extreme market conditions and high levels of fear in the market. Extreme market conditions also increase the amount of noise and it has attracted a higher number of retail investors in the market. To conclude from the introduced theories, this thesis seeks to shed light if VIX is an explanatory force for retail investor activity, relating to weak form of efficiency and VIX includes information of market wide retail investor activity. Further so, herding effect as a combination of noise trading and positive feedback trading could suggest that if level of VIX is the level of fear, short term decrease of VIX could translate to increased retail investor activity, as the sentiment turns to more positive. Though, in the context of retail investors, VIX might be more of a signal of noise, suggesting that it is significant signal for retail investor activity only at higher levels, and in that case, short term upswings should translate to increase in retail investor activity.

2.7 Conceptual framework

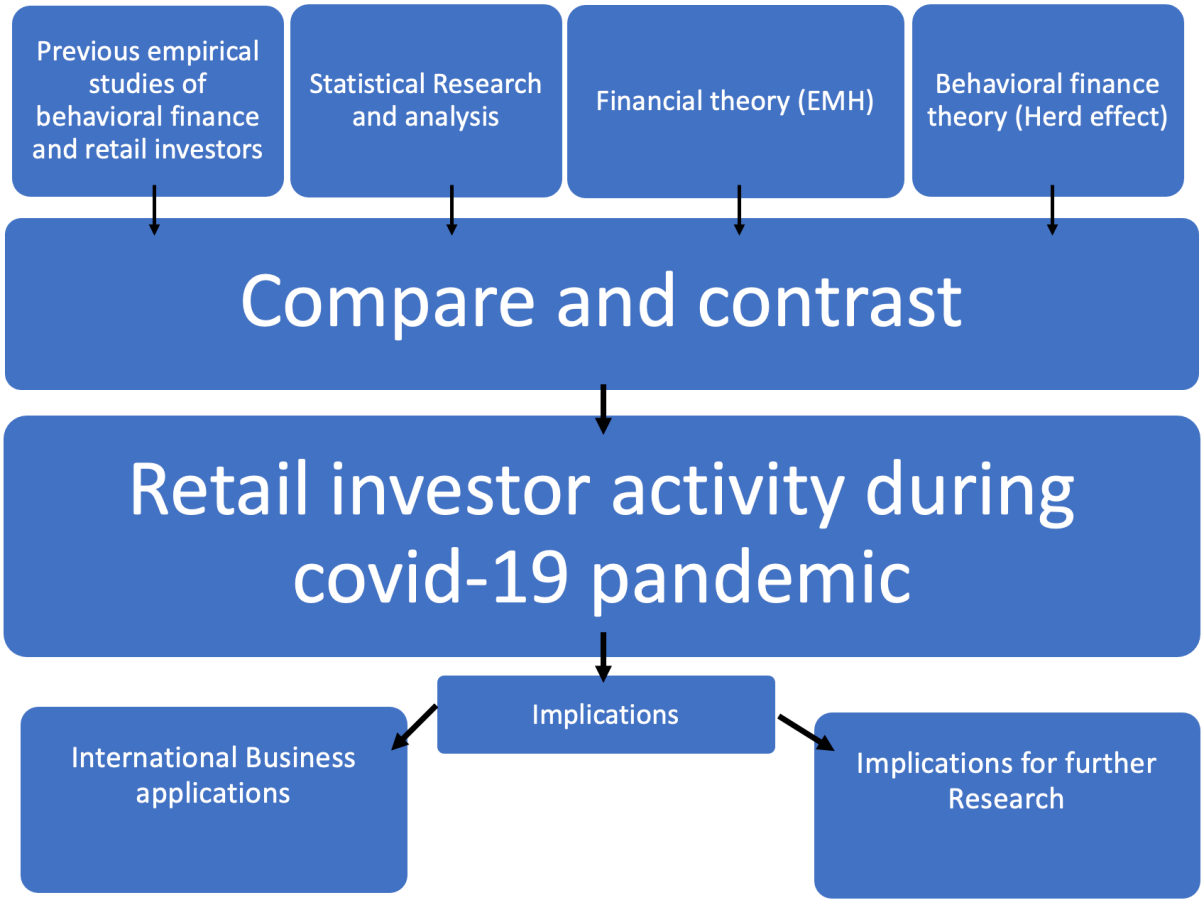

Figure 1: Conceptual Framework (Uusi-Mikkola, 2021)

2.7.1 Hypothesis

On the basis of the literature review we have established that retail investor activity has increased during the pandemic, retail investors behave systematically, but on irrational signals and market volatility on top of its obvious measure, can be a proxy for noise and market fear. Drawing from the research objective of the thesis, the goal is to understand how retail investor experience established market indicators and does it have explanatory power over retail investor activity. The timeframe for this thesis includes data before and during the covid-19 pandemic. The hypotheses created for the research, separated to corresponding research questions, are as follows:

1. Does VIX provide predicting power over retail investor activity?

H1: VIX contains information of retail investor activity, and retail investor activity is correlated with VIX.

2. How does the level of VIX affect retail investor activity?

H2: During higher levels of VIX, the overall retail investor activity increases, having more economical power and being more predictable.

3. How retail investors account for variation in VIX under different market conditions?

H3a: If VIX is a proxy for retail investor fear, absolute negative change in VIX should correlate with the aggregated total holdings of Robinhood.

H3b: If VIX is a proxy for noise, change in VIX should correlate with the aggregated total holdings of Robinhood.

H1 is tested via null hypothesis of the overall regression model. The null hypothesis is $H0: \alpha = 0$, has two alternate hypotheses if $H1: \alpha \neq 0$, we reject the null hypothesis. The null hypothesis is at confidence level of 95% and represented in the Excel regression output as *significance F,* or as p-value of the F-statistic, which translates to the statistical significance of the regression model. The null hypothesis is tested for the three different time frames, the full timeframe from May 2018 to August 2020, the time before the pandemic (May 2018- end of 2019) and during the pandemic (2020 until 13th of August).

H2 is tested by comparing the different timeframes and the statistical significance between the overall models by assessing the robustness of the model and the significance indicators, such as R-squared, F statistic, and standard error.

H3 is tested by investigating the statistical significance of each independent variable at each timeframe. If the p-value of an individual variable of a model that conforms to OLS assumptions, falls within the 95% confidence (p-value<0,05), it is considered significant.

3 METHODOLOGY

The methodological approach taken in this thesis seeks to answer the research problems and test the hypotheses with a primary quantitative analysis of secondary data. The methodologies used in the thesis are inspired by those used in other studies of VIX and investor behavior: Welch (2020), Lei et. al (2012), Foucault et. al (2011) and Aharon & Quadan (2020). This thesis conducts research to shed a light if VIX

index could be used to create a predictive model for retail investor activity in the market.

Firstly, the analysis focuses on descriptive data. This section explains the datasets, the sample and gives an overview of each set to provide better understanding of the sample, followed with a graph-based analysis visually describing the datasets. In the second part of the analysis is the inferential statistics, which provides deeper understanding and tests the introduced hypotheses via regression analysis. The regression analysis is divided into three parts, with pre-pandemic, pandemic and the full timeframe.

3.1 Data

3.1.1 Robintrack data

The retail investor data is collected from Robintrack.com, which collected popularity data from Robinhood's API from May 2018 until 13th of August 2020, the dataset includes hourly (or minutely, slight variance on some days) data of each security traded on Robinhood, altogether 8597 separate files for each ticker. These individual csv files were combined into one master file and to limit the size, for each day for each ticker, the max value was extracted, this resulted into 818 rows (days) of usable data. Either Robintrack or Robinhood experienced couple breaks without data collection, which were removed from the dataset (17 days were removed in total). The popularity data is the amount of Robinhood users holding a stock at a specific time.

Robintrack data offers a barrier free access to retail investor data. The shortcomings of this dataset are the bias in the user group, as the barrier-free platform might attract beginners and not be perfect representation of retail investors, and popularity data only shows number of users holding a stock and not the absolute number of stocks held on the app, nor the actual trades. The approximations of retail investor behavior based on this dataset, are the changes in the aggregated total holdings. Also, the data can be assessed as the ratio of how highly is the Robinhood crowd invested in the market (Welch, 2020).

The total amount in the end of the sample is over 42 million, and in the beginning if the sample it is approximately 5.5 million, this ratio (13% in the beginning of the sample) represents the used capital of the available capital from the Robinhood investors, or in other words, the amount of available assets were 13% stocks and 87% cash at that point, whereas when the popularity reached its peak in 10th of August 2020 100% of the available assets were on stocks.

Major short coming with the dataset is that new users bias the dataset, as Robinhood has offers each new user a one free share, which then is displayed on the data. As the Robintrack popularity data merely shows the amount held at the specific time, the proxy for retail investor activity is the aggregated logarithmic daily change of the total number of users holding securities on Robinhood.

3.1.2 CBOE VIX-data

CBOE offers historical, and current data downloads for free on their Website, this is beneficial for research use and it is one reason for the wide use of VIX for academic and research purposes. The VIX data includes daily Open, Low, High and Close values for VIX for every day NYSE is open, which totals to 564 usable observations after removing the days Robintrack did not provide data. VIX is used as a proxy for market sentiment in the context of this study, as well as explanatory variable for retail investor behavior.

High values of VIX translate to higher fear in the market, thus lower market sentiment, and possibly higher noise. In the chart-based analysis the timeseries chart of historical values of VIX and the daily return of VIX are assessed. The direction of change in VIX tells short term effect on retail investor behavior, the level of VIX illustrates the threshold, when VIX is relevant to investors, both of which are analyzed in the research.

3.2 Analysis

3.2.1 Chart based analysis

The first part of analysis is to investigate the data and the movements of the indices and the popularity data during the timeframe 2nd of May 2018 to 13th of August 2020, which includes pre-pandemic era and the time during the pandemic in the year 2020. Visual analysis provides initial insight of the datasets and possible explanatory power of the two datasets.

First, both datasets are analyzed based on the raw values, and secondly the transformed graphs. Both datasets were transformed to logarithmic daily return: $r_t = LN(\frac{v(t)}{v(t-1)})$

where r_t is the logged average daily return of the dataset, and $v(t)$ is the value at the time of examination and $v(t-1)$ is the previous value. The logarithmic transformation improves the fitness for the regression analysis.

3.2.2 Regression analysis

The regression analysis is the pinpoint of this thesis. Welch (2020) analyzed logged daily change of Robinhood investors in context of logged change of daily returns of S&P 500 in time series regression. The regression analysis in this thesis replicates that methodology with substituting S&P500 returns with changes of VIX. The function of the multilinear regression in this thesis is similar, to assess how the variance in independent variables affect the dependent variable.

The hypotheses are tested on the regression analysis of the two variables introduced in the section 3.1, which are then transformed to separate variables for positive and negative change of VIX. The regression is as follows:

$$Log\Delta RT_t = b_0 + b_{VIXp} \times \Delta VIXp_t + \cdots + b_{VIXnlag5} \times \Delta VIXnlag5_t$$

Where:

$\boldsymbol{Log\Delta RT_t}$: The logarithmic change of RT at time t

$\boldsymbol{b_0}$: The intercept

$\boldsymbol{b_{VIXp}}$: the regression coefficient of positive change of VIX

$\Delta \boldsymbol{VIXp_t}$: the positive change of VIX at time t

$\boldsymbol{b_{VIXnlag5}}$: The regression coefficient of 5-day lagged negative change of VIX

$\Delta \boldsymbol{VIXnlag5_t}$: The 5-day lagged negative change of VIX at t

The VIX is divided into 12 independent variables, for positive change of VIX and accordingly 1-, 2-, 3-, 4- and 5-day lags, and the same for the negative change of VIX. Additionally, the negative changes at VIX are transformed to absolute values for easier interpretation thus, positive coefficient with either of the variables translates as positive correlation in the dependent variable. Positive (negative) change variable returns a value of 0 on a day of negative (positive) change. Logarithmic change of VIX is divided into positive change and negative change to see if the direction of change affects retail investor behavior differently, this is crucial for testing the hypothesis H3 for understanding the rationale behind the activity. The separated lagged data helps illustrate how long a change in either direction in VIX takes to have effect on the aggregated total popularity of RH holdings.

The fitness for the regression analysis is assessed via visual analysis of the variables. The assumptions are those of Markov-Gauss seven assumptions:

1. Linearity of the variables
2. The error term has a population mean of 0
3. Independent variables are uncorrelated with the error term
4. Observated error terms are uncorrelated with each other
5. Homoscedasticity, error term has constant variance
6. Independent variables are not perfect linear functions of other explanatory variables
7. The error term is normally distributed.

When these assumptions are held true with the regression analysis, the model provides robust results. However, in many cases empirical data fails to completely fulfill these assumptions, which causes reliability issues to the model. Also, variety of statistical tools exist to mitigate the issues meeting the assumptions with a regression model.

To understand the credibility and the findings of the regression analysis, the assumptions need to be assessed. The Markov-Gauss assumptions are visually analyzed based on plots of the datasets and the regression outputs for each timeframe separately.

4 FINDINGS

4.1 Chart based analysis

4.1.1 Robintrack data

Figure 2 plots the aggregates total number of holdings during the timeframe under examination on Robinhood, in other words the sum of all the holdings in timeseries. From the plot can be seen constant growth until early 2020, or the beginning of the Covid-19 pandemic, when Robinhood experienced a significant growth of userbase. The plot illustrates the increased interest of retail investors into stock markets during the pandemic. Figure 3 plots the logged change of users holding securities and additionally the 7-day moving average to better illustrate the movement. From Figure 3, can be seen that the changes in holdings increased significantly in the year 2020, the sharp spike in January 2020 can be explained with a system outage, causing a one-day spike.

Comparing the moving average to the chart of VIX values, can we see that during higher levels of VIX the changes in total holdings were more positive, and higher than previously.

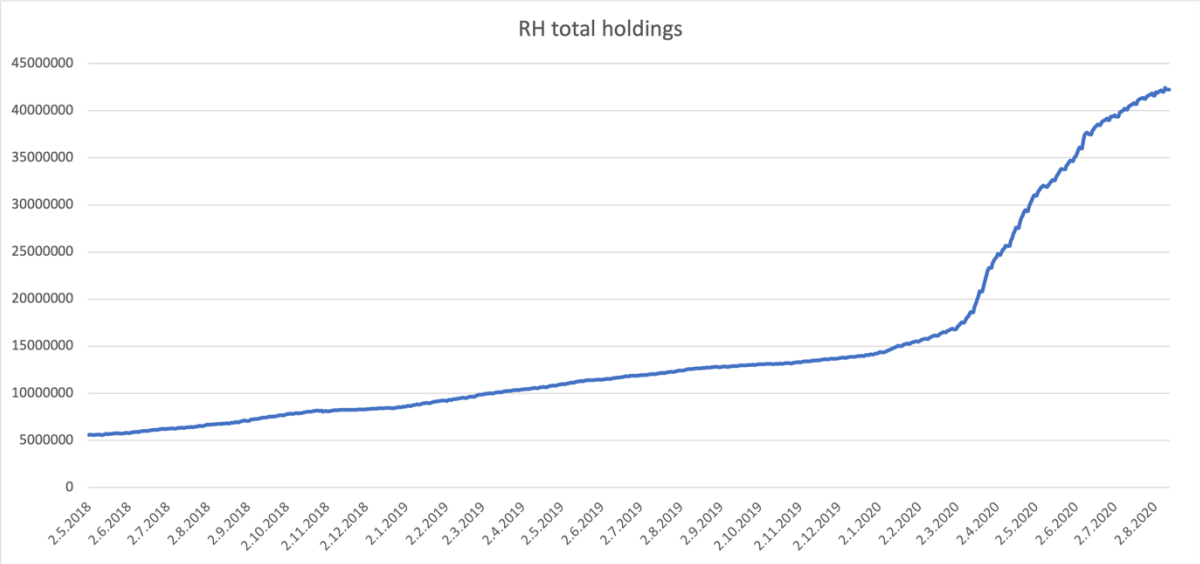

Figure 2 – Aggregated total number of users holding stocks on Robinhood

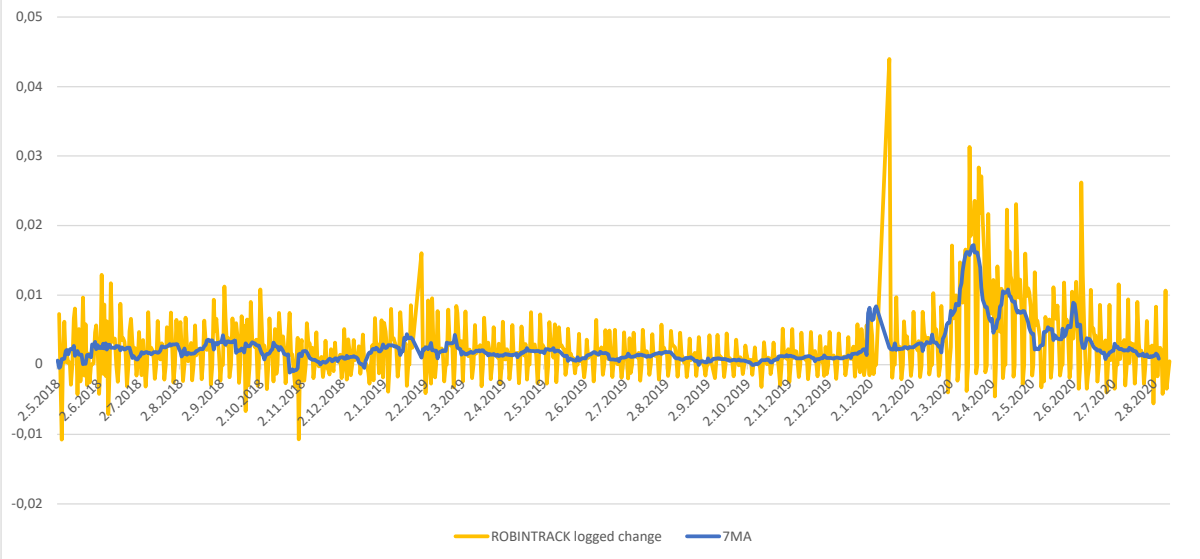

Figure 3: Logarithmic daily change of aggregated total holdings and 7-day moving average

4.1.2 VIX- data

Figure 4 plots the timeseries of closing values of VIX. During the timeframe under study, VIX was relatively low from May 2018 until January 2020. In February 2020 VIX soared to extreme values to over 80 and managed to come down to near 20 by August of 2020. This timeseries is used to divide the timeframe into two parts for the regression analysis; the pre-pandemic time with lower VIX from May 2018 to end of December 2019 and the Pandemic with higher values of VIX from January 2020 to mid-August 2020. Figure 5 plots the logged daily change of VIX.

From raw visual comparison between the timeseries of values and the logged change of VIX, slight increases in the value cause large variance in the logged changes. During the higher levels of VIX the variance quickly returns after the spike in March 2020.

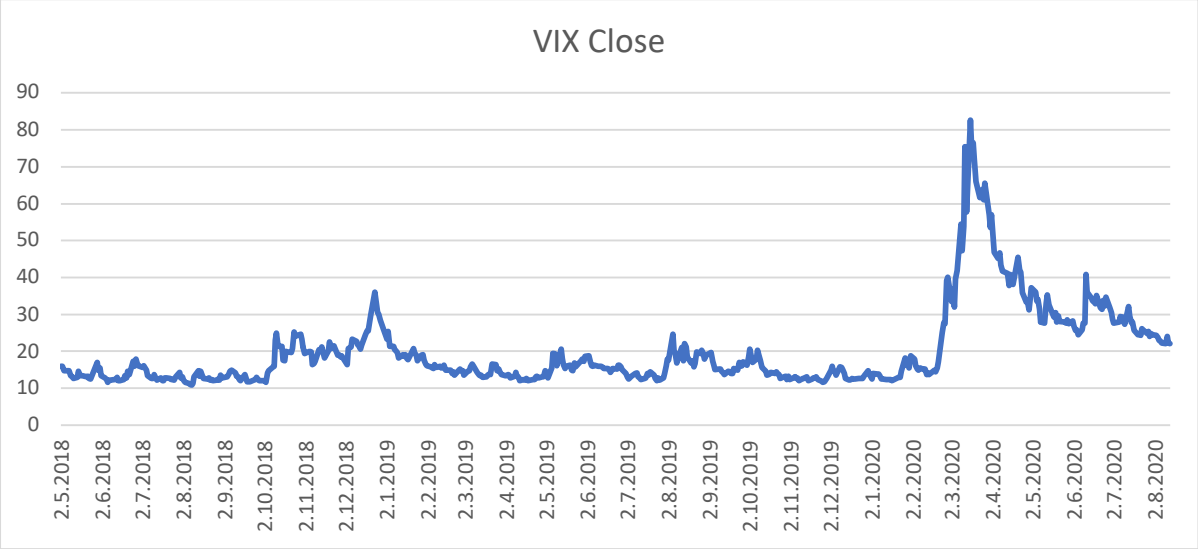

Figure 4: Raw values of VIX

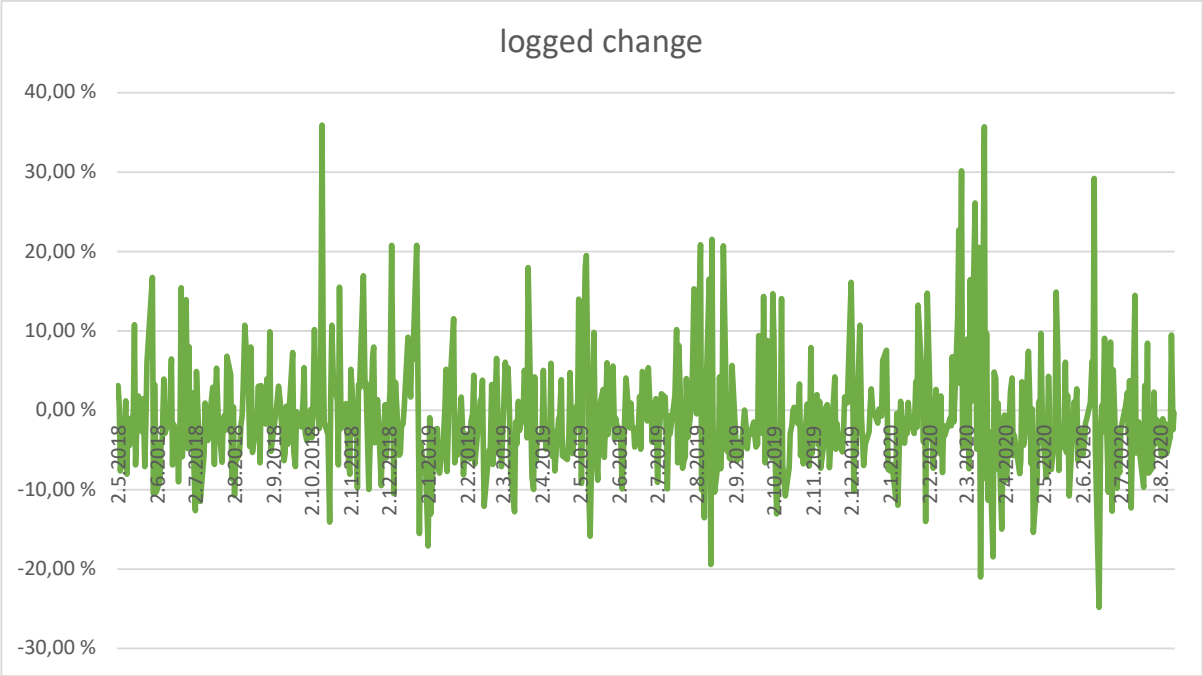

Figure 5: Logarithmic daily change of VIX

4.2 Regression analysis

Multilinear regression analysis is paramount in this thesis, as these are the only tests for the hypotheses. The null hypotheses are tested for three timeframes, first the full timeframe of the sample, then the pre-pandemic and lastly the timeframe during the

pandemic. For each timeframe, the OLS assumptions are assessed visually to understand the value of the regression output. The linear regression is formed from the independent variable, logged change of VIX, and the dependent variable, logged change of aggregated total of RH users holding. Both variables are logged for better fitness for OLS analysis.

From visual standpoint, much cannot be deducted from the two variables, as both of them experience intense variance during the studied timeframes. The key variable tested in the regression is the β of the various independent variables, or as it is in the regression output from Excel, the coefficient, which explains the direction of possible correlation and thus, further explains the explanatory power of the variable. However, to be able to deduce anything from the coefficient, the statistical significance of the results needs to be examined. The R squared with the standard error gives goodness of fitness of the regression model, and the P-values represent the significance of the independent variables. All of the examined values are highlighted in the results.

4.2.1 Full timeframe

Figure 2 plots the predicted values against the observed values of the regression model in the full timeframe to observe the linearity of the model. From the basis of the chart, the linearity is not exactly strong, but somewhat existing. This forbids this models' use for extrapolating any future events, but rather shows that it may still be used to assess, if the variance of the independent variables may explain variance of the dependent variable to some degree and to analyze the existence of this relationship. The results of the separated timeframes provide better understanding of the full timeframe.

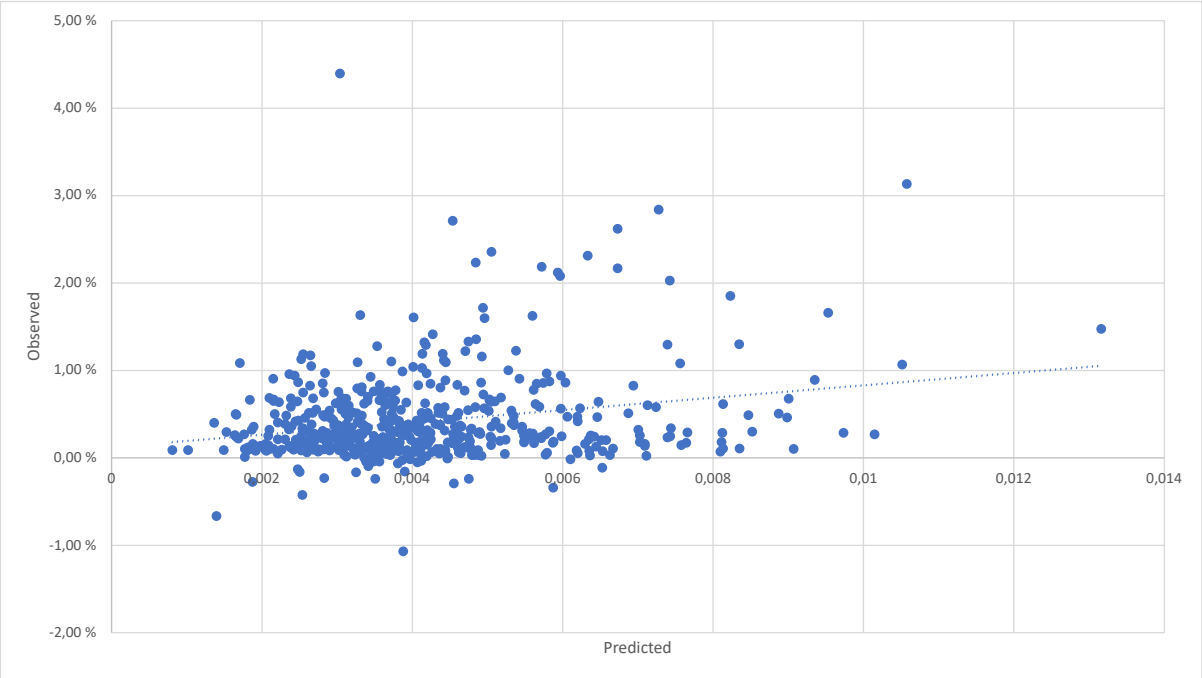

Figure 6 – Scatter plot of observed and predicted results

Figure 7 plots the residuals of the regression. To assess the second assumptions, the mean of the observed error term should be 0 and based on the chart, the mean of the observed errors is nearly 0. This thesis applies timeseries data to OLS, and by nature, time series data is problematic in terms of autocorrelation, and we may assume that the variables are autocorrelated. Additionally, assessing the scatterplot of the residuals, the increase in variation in the values suggests heteroscedasticity. Thus, to account for heteroscedasticity and autocorrelation, the regression analysis applies Newey-West standard errors and t-statistics of 2-lags, as done by Welch (2020), to improve the robustness of the results.

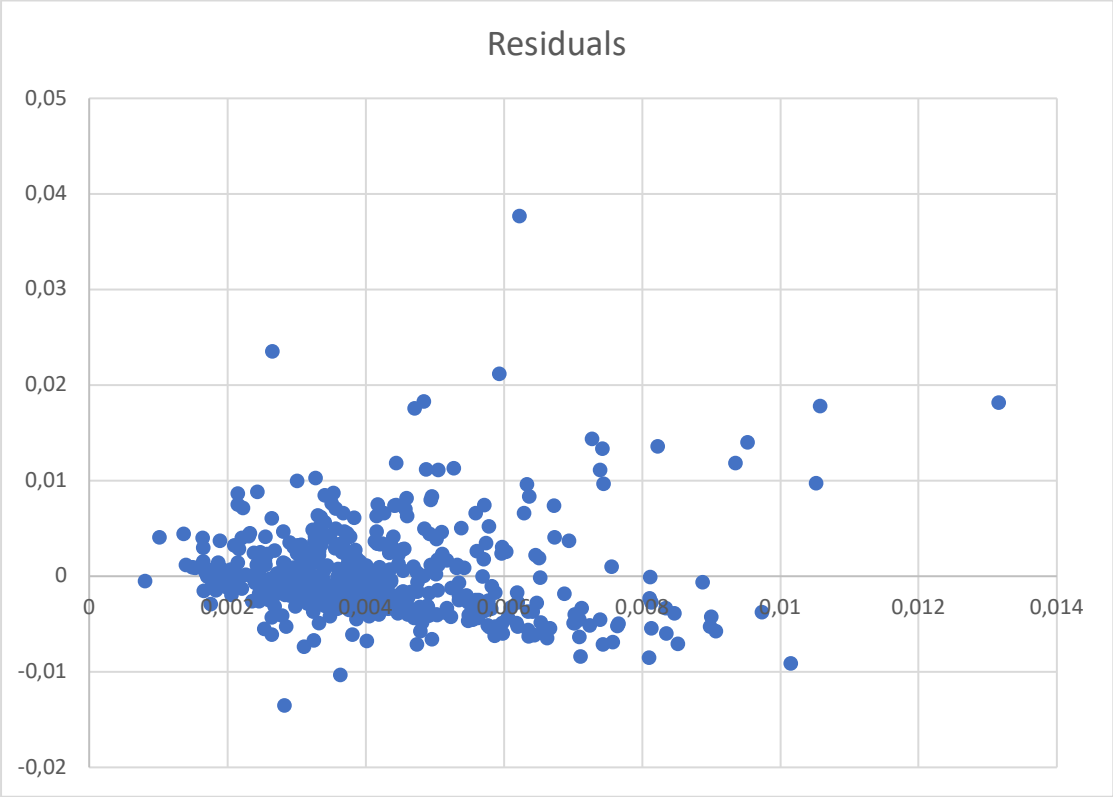

Figure 7 – Scatter plot of the residuals

Table 1 shows the results output of Excel of the OLS regression for the full timeframe, from 9th of May 2018 to 12th of August 2020, number of observations totaling to 554 after matching the data between the datasets. The F statistic is rather low at 5.8 however, the significance F is well within the 95% significance. The R squared has a value of 0,11 which suggests that the model only is attributable for 11% change in total holdings during the time period. This finding is in line with the observations of figure 6 which shows that the model does not fit the sample very well in the full timeframe.

Reviewing the the P-values, VIXn, VIXnlag1, -3, -4 and -5 fall within the 5% significance and VIXnlag2 is at 10% significance. VIXp and VIXplag5 are also within the 5% significance level. These suggest that in the full timeframe, retail investors are more likely to increase their position when VIX experiences downturns, but with a slight delay. The VIXp results suggest that retail investors act quicker to increases in VIX, but with less magnitude. The null hypothesis for the regressions is: $H0: b = 0$. The null hypothesis is rejected based on the values of the regression output, but due to the low linearity of the model, there is a risk of type I error.

OVERALL FIT							
Multiple R	0,33638124		AIC	-5982,1974			
R Square	0,11315234		AICc	-5981,42534			
Adjusted R Sc	0,09366118		SBC	-5925,95746			
Standard Erro	0,00469014						
Observations	559						
ANOVA				Alpha	0,05		
	df	*SS*	*MS*	*F*	*p-value*	*sig*	
Regression	12	0,00153242	0,0001277	5,80531665	1,6642E-09	yes	
Residual	546	0,01201058	2,1997E-05				
Total	558	0,013543					
ols	*coeff*	*std err*	*t stat*	*p-value*	*lower*	*upper*	*vif*
Intercept	0,00037759	0,00053737	0,70267202	0,48255964	-0,00067797	0,00143315	
VIX-	0,01312219	0,0059255	2,21452991	0,02720453	0,00148263	0,02476176	1,33910913
lag1VIX-	0,01825485	0,00588648	3,10114638	0,00202727	0,00669192	0,02981778	1,32297626
lag2VIX-	0,01086645	0,00580405	1,87222033	0,06171042	-0,00053454	0,02226745	1,28618006
lag3VIX-	0,01639099	0,00569974	2,87574194	0,00418815	0,00519488	0,0275871	1,24178394
lag4VIX-	0,01141746	0,00564156	2,02381053	0,04347617	0,00033563	0,02249929	1,21716077
lag5VIX-	0,02002273	0,00557146	3,59380297	0,00035537	0,00907861	0,03096685	1,18815884
VIX+	0,01032755	0,00444114	2,32542833	0,02041455	0,00160374	0,01905137	1,28948832
lag1VIX+	0,00314681	0,00439069	0,71670101	0,47386505	-0,0054779	0,01177151	1,26035569
lag2VIX+	0,00622862	0,00446197	1,39593595	0,16330112	-0,00253611	0,01499335	1,29752605
lag3VIX+	0,00689412	0,00450814	1,52925971	0,12677919	-0,00196131	0,01574955	1,32451886
lag4VIX+	0,00128127	0,00453334	0,28263228	0,77756585	-0,00762365	0,01018618	1,33936336
lag5VIX+	0,01210598	0,00461683	2,62213988	0,00898182	0,00303705	0,02117491	1,3891564
newey-west	*coeff*	*std err*	*t stat*	*p-value*			
Intercept	0,00037759	0,00102531	0,36827253	0,71281277			
VIX-	0,01312219	0,00724494	1,81122145	0,07065597			
lag1VIX-	0,01825485	0,00874035	2,0885718	0,03720889			
lag2VIX-	0,01086645	0,0075773	1,43408068	0,15212169			
lag3VIX-	0,01639099	0,00685265	2,39191959	0,0170978			
lag4VIX-	0,01141746	0,00667513	1,71044754	0,0877512			
lag5VIX-	0,02002273	0,00751534	2,66424923	0,00794411			
VIX+	0,01032755	0,00614941	1,67943714	0,09363896			
lag1VIX+	0,00314681	0,00572825	0,54934867	0,58299079			
lag2VIX+	0,00622862	0,00532447	1,16981094	0,24258745			
lag3VIX+	0,00689412	0,00513899	1,34153314	0,18030511			
lag4VIX+	0,00128127	0,00576789	0,22213789	0,82428961			
lag5VIX+	0,01210598	0,00685661	1,76559318	0,07802299			

Table 1: Regression output of the full timeframe

4.2.2 Pre-pandemic

Figure 8 plots the predicted values and the observed values in the chart, which shows non-linearity within this timeframe. This forbids the hypothesis testing in this timeframe as the model provides no significance and the variables do not correlate during this timeframe.

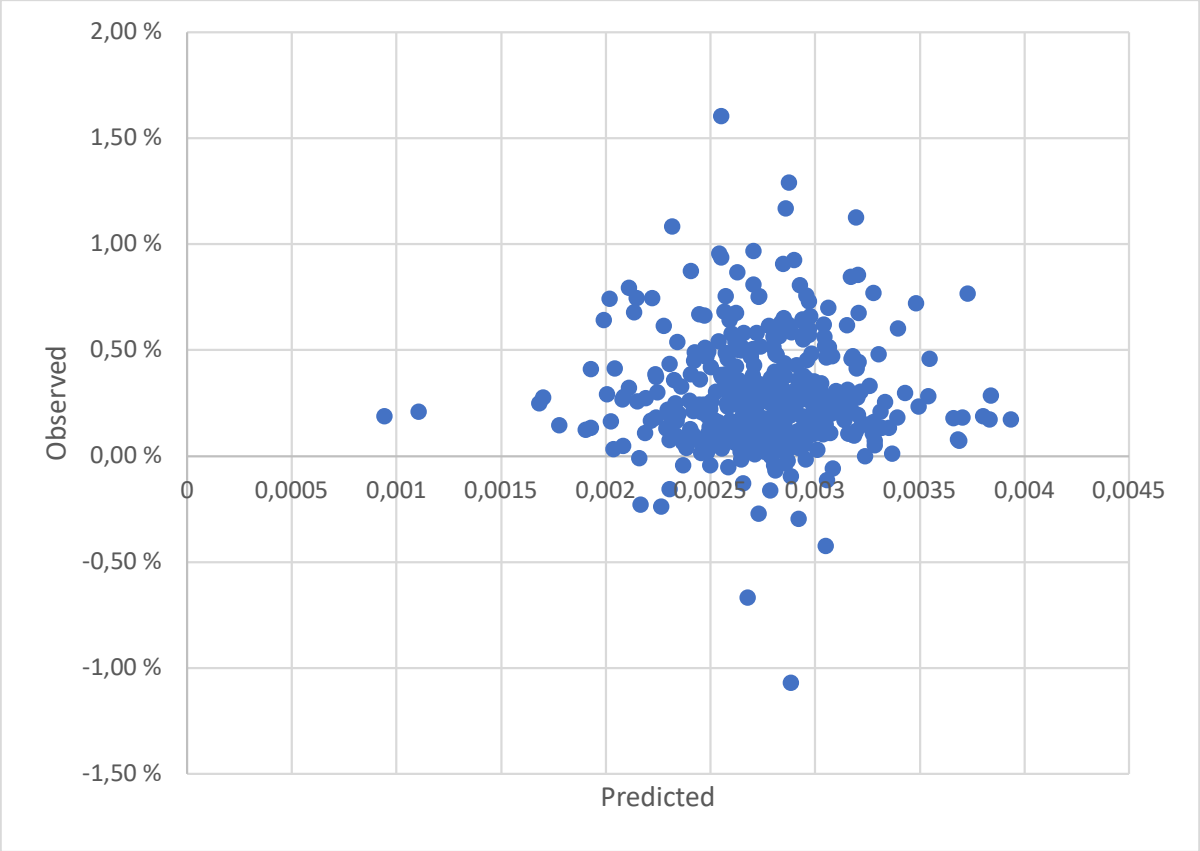

Figure 8 – Pre-pandemic predicted values and observed values

Table 2 shows the results of the second regression, which includes the timeframe from 9th of May 2018 until 31st of December 2019. The pre-pandemic results are interesting compared to the full timeframe, as no statistical significance exists. The significance F is almost at 0,8 which fails to fall within the 95% significance, also R-squared is at 0,019% meaning that the model shows no statistical significance. None of the P-values of the independent variables show statistical significance either. The results of this timeframe undermine the significance of the full timeframe, as this displays that most of the observations in the first timeframe provide no statistical significance.

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0,14004208							
R Square	0,01961178							
Adjusted R S	-0,0103237							
Standard Err	0,00267846							
Observations	406							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	12	5,64E-05	4,7E-06	0,65513428	0,79423682			
Residual	393	0,00281944	7,1741E-06					
Total	405	0,00287584						
	Coefficients	*tandard Erro*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95,0%*	*Upper 95,0%*
Intercept	0,00270049	0,00036166	7,46684734	5,351E-13	0,00198946	0,00341153	0,00198946	0,00341153
VIXn	0,00485377	0,00431884	1,12385885	0,26175892	-0,0036372	0,0133447	-0,0036372	0,0133447
VIXnlag1	0,00352294	0,00430297	0,81872213	0,41344113	-0,0049368	0,01198265	-0,0049368	0,01198265
VIXnlag2	-0,0039429	0,00423884	-0,9301947	0,35284131	-0,0122766	0,00439069	-0,0122766	0,00439069
VIXnlag3	0,00458592	0,00412932	1,11057409	0,26743074	-0,0035324	0,01270425	-0,0035324	0,01270425
VIXnlag4	-0,002824	0,00412731	-0,6842167	0,4942416	-0,0109383	0,00529039	-0,0109383	0,00529039
VIXnlag5	0,00028987	0,00408741	0,07091801	0,94349908	-0,0077461	0,0083258	-0,0077461	0,0083258
VIXp	0,00020962	0,0032292	0,06491362	0,94827578	-0,0061391	0,00655829	-0,0061391	0,00655829
VIXplag1	-0,0039733	0,00324655	-1,2238544	0,22174039	-0,0103561	0,00240947	-0,0103561	0,00240947
VIXplag2	-0,0020916	0,00325563	-0,6424412	0,52096134	-0,0084922	0,00430908	-0,0084922	0,00430908
VIXplag3	0,00304066	0,00333317	0,91224481	0,3621991	-0,0035124	0,00959373	-0,0035124	0,00959373
VIXplag4	-0,0040717	0,00338267	-1,2036998	0,22943016	-0,0107221	0,00257867	-0,0107221	0,00257867
VIXplag5	0,00093458	0,00341077	0,2740082	0,78422228	-0,0057711	0,00764022	-0,0057711	0,00764022

Table 2: Regression output of the Pre-pandemic timeframe

4.2.3 Pandemic

The third part of the regression analysis fits the OLS assumptions better than the previous two, showing strongest linear relationship, shown on figure 9. This undermines further the significance of full time-frame findings, showing that the small correlations are due to the inclusion of pandemic in the timeframe.

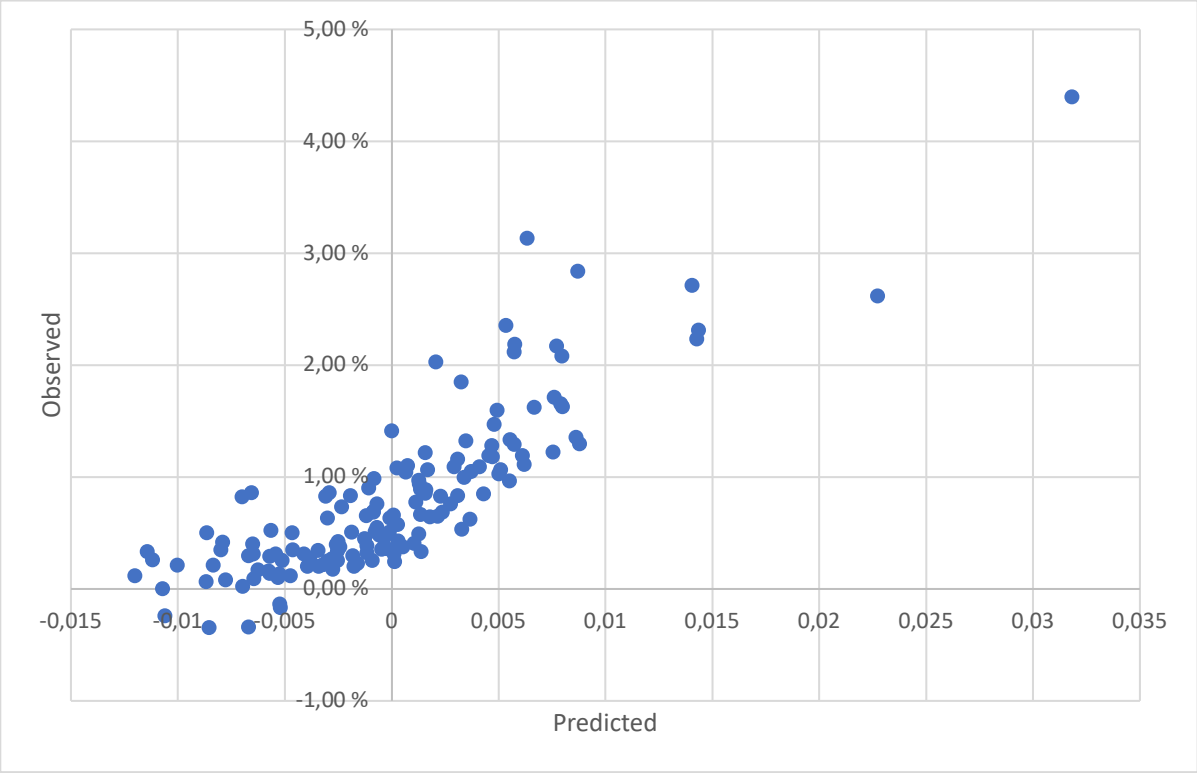

Figure 9 – Pandemic predicted values and observed values

The residual output plotted in figure 10, shows that the error term has approximately a mean of 0 and is randomly distributed, however heteroscedasticity exists, and we may assume some level of autocorrelation to exist within the timeseries data, thus again applying the Newey-West standard errors and t-statistics to account for these weaknesses in the model.

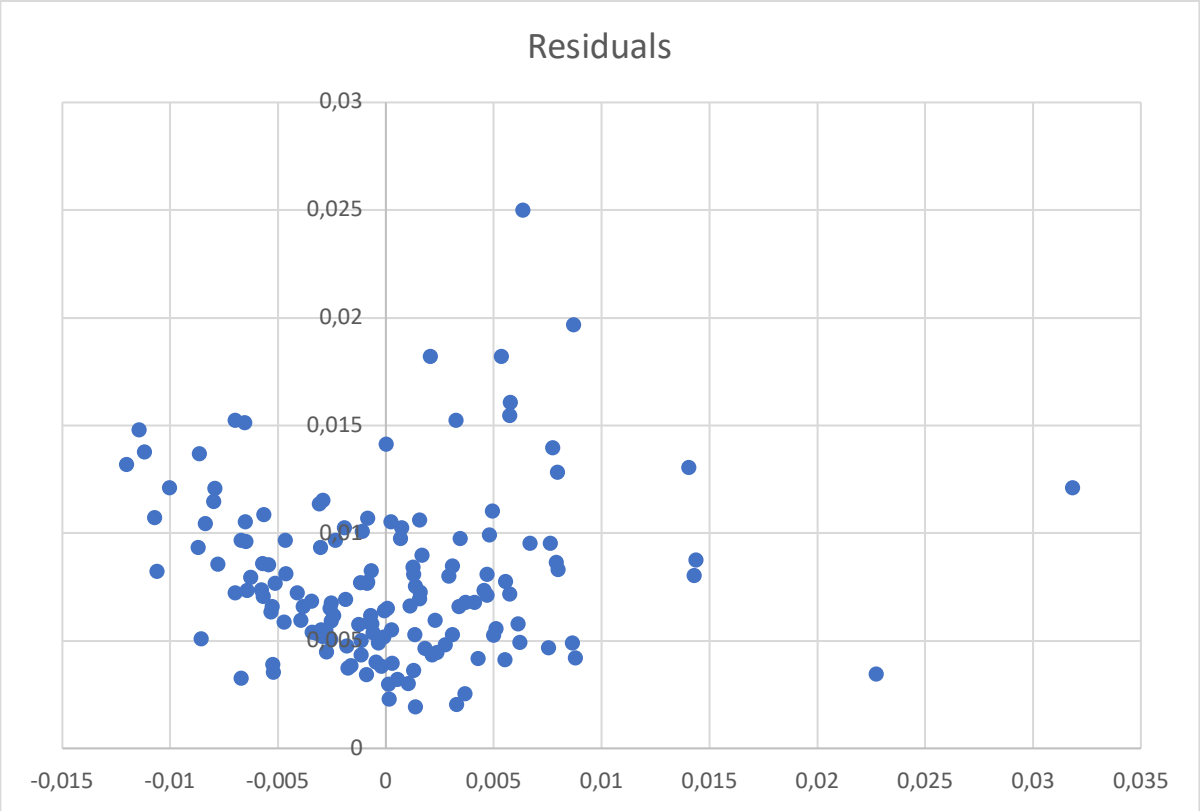

Figure 10 – Pandemic residual output

Table 3 shows the results of the third and final regression, which gives no surprises, after the first two timeframes. As could be deducted from the two previous regressions, during the pandemic, the model shows significance between the variables. The p-value of the F statistic shows a statistical significance at 95% confidence. The same independent variables have p-values at significant level as with the full timeframe regression, with the difference of VIXnlag2, which is also at 5% significance level. Additionally, the confidence levels of each variable improve the reliability of the results as each of them remain positive in the full 95% confidence level. The results illustrate that retail investors act to negative change in VIX more, but with a delay and to positive change of VIX quicker, but with also a 5-day delay, suggesting momentum trading behavior, similar to evidence Foucault et. al (2011) mentioned with the difference of idiosyncratic volatility, rather than an index. The results from the regression show that during the pandemic VIX shows better correlation to retail investor behavior than during “regular market conditions”, though by itself this regression model has very low forecasting value, as the R squared is only at 27%.

OVERALL FIT							
Multiple R	0,5241006		AIC	-1480,8328			
R Square	0,2746815		AICc	-1477,6749			
Adjusted R S	0,2102087		SBC	-1441,869			
Standard Er	0,0064435						
Observation	148						
ANOVA				Alpha	0,05		
	df	*SS*	*MS*	*F*	*p-value*	*sig*	
Regression	12	0,0021227	0,0001769	4,2604272	1,07E-05	yes	
Residual	135	0,005605	4,152E-05				
Total	147	0,0077277					
ols	*coeff*	*std err*	*t stat*	*p-value*	*lower*	*upper*	*vif*
Intercept	-0,0022645	0,0017022	-1,3302907	0,1856654	-0,005631	0,001102	
VIX-	0,028499	0,0137609	2,0710083	0,0402626	0,0012841	0,0557138	1,3699872
lag1VIX-	0,045811	0,0136962	3,3447864	0,0010663	0,018724	0,0728979	1,3630586
lag2VIX-	0,0383824	0,0135703	2,8284186	0,0053912	0,0115446	0,0652203	1,3381052
lag3VIX-	0,030947	0,0135176	2,2893816	0,0236089	0,0042133	0,0576806	1,333416
lag4VIX-	0,0341966	0,0132185	2,5870348	0,0107385	0,0080546	0,0603387	1,2770563
lag5VIX-	0,047633	0,0128696	3,7012015	0,0003115	0,0221809	0,073085	1,2141492
VIX+	0,0219011	0,0106446	2,0574768	0,0415651	0,0008493	0,0429529	1,5689567
lag1VIX+	0,0109516	0,0100883	1,0855787	0,2796001	-0,0089999	0,0309032	1,4092422
lag2VIX+	0,0113409	0,010499	1,080182	0,2819871	-0,009423	0,0321048	1,5205101
lag3VIX+	0,003486	0,0104517	0,333537	0,7392464	-0,0171843	0,0241564	1,5076364
lag4VIX+	0,0022255	0,0101368	0,2195428	0,8265589	-0,017822	0,0222729	1,4181428
lag5VIX+	0,0218941	0,0105545	2,0743886	0,0399428	0,0010206	0,0427677	1,5371861
newey-west	*coeff*	*std err*	*t stat*	*p-value*			
Intercept	-0,0022645	0,0023123	-0,9793015	0,3291827			
VIX-	0,028499	0,0135376	2,1051758	0,0371287			
lag1VIX-	0,045811	0,0164664	2,7820868	0,0061748			
lag2VIX-	0,0383824	0,013253	2,8961415	0,0044084			
lag3VIX-	0,030947	0,0130939	2,3634684	0,0195309			
lag4VIX-	0,0341966	0,0128412	2,6630374	0,0086857			
lag5VIX-	0,047633	0,0128264	3,713654	0,000298			
VIX+	0,0219011	0,008651	2,5316358	0,0124993			
lag1VIX+	0,0109516	0,0093846	1,1669791	0,245276			
lag2VIX+	0,0113409	0,0074371	1,5249046	0,1296222			
lag3VIX+	0,003486	0,0083706	0,4164635	0,6777323			
lag4VIX+	0,0022255	0,007205	0,3088761	0,757892			
lag5VIX+	0,0218941	0,0099777	2,1943062	0,0299238			

Table 3: Regression output of the Pandemic timeframe

The hypotheses introduced were as follows:

H1: H1: VIX contains information of retail investor activity, and retail investor activity is correlated with VIX.

H1 is proven when the null hypothesis of the pandemic timeframe is rejected, which shows statistical significance, thus showing that during the pandemic, VIX had explanatory information of retail investor activity.

H2: During higher levels of VIX, the overall retail investor activity increases, having more economical power and being more predictable.

H2 is proven as the pandemic timeframe showed statistical significance, whereas the pre-pandemic timeframe failed to fulfill any statistical significance and due to inclusion of the pre-pandemic in the full timeframe, the pandemic timeframe is the

only one having statistical significance, ultimately proving the H2 in the scope of this study.

H3a: If VIX is a proxy for retail investor fear, absolute negative change in VIX should correlate with the aggregated total holdings of Robinhood.

H3b: If VIX is a proxy for noise, positive change in VIX should correlate with the aggregated total holdings of Robinhood.

Both alternatives of Hypothesis 3 are partly confirmed, as the behavior of retail investors differs between the direction of change of VIX. Both directions have strongest correlation after 5 days, suggesting that powerful move in any direction in VIX results in increase in retail investor activity after 5 days. The results are more in favor of VIX being a proxy for noise, as the VIX needs to be at high enough level to give any predictive power. However, at higher level, it could be a proxy of fear, as after a downturn, retail investors increase their positions – in accordance to improving market sentiment.

5 DISCUSSION AND ANALYSIS

This thesis provides further evidence of tools for following retail investors in the market. The findings are on par with previous studies such as: Boehmer et. al, (2016); Ekkayokkaya et. al, (2020); Moss et. al, (2020) and Pagano et. al, (2020), who found evidence of increase in economic significance of retail investors during more extreme market conditions. However, the R square values even during the pandemic were low, as well as the coefficients, showing that VIX only has additional explanatory power to retail investor behavior and combining it with data from market returns or other explaining variables could provide more robust model for forecasting retail investor activity. The findings support the results of previous studies showing that during the covid-19 pandemic, retail investors provided liquidity to the market with liquidity with contrarian investing strategies, as Lei et. al (2012) showed positive correlation with VIX and trading volume, would this further show evidence of retail investors behaving in the market as stabilizing force, as the total popularity increased after downturn in VIX, which should translate to downturn in trading volume. Also, the correlation with

positive turns in VIX shows evidence of noise trading and momentum seeking, as the effects were significant at the day of the positive VIX and five days afterwards, which could be due to the time of a bank transfer, similar effect that was recognized by Welch (2020) in context of price changes of S&P 500. On the other hand, during the pandemic after a downturn in VIX, investors were more attracted to the market, suggesting that during high levels of VIX, a negative change reduces fear in the market and retail investors increase their position for five following days.

Aharon and Quadan (2020) showed correlation with high changes in VIX and attention to trading platforms. The findings of this thesis are supporting these results as the attention to platform could realize into increased activity in the market with a delay. The null hypothesis $H0: \alpha = 0$ was rejected only at full timeframe and during the pandemic. Despite the rejected null hypothesis at full timeframe, the explanatory value of VIX to retail investor behavior is only applicable during higher levels of VIX, or extreme market conditions, this is linked to the OLS assumptions as the pandemic timeframe fulfilled those the best out of the three timeframes.

The findings point towards VIX being more of a source of noise than fear with retail investors contradicting the thoughts of Lei et. al (2012) using VIX as a proxy for investor sentiment, though retail investors as a group might not fit the established thoughts of overall market in terms of sentiment. After a certain threshold of VIX, retail investors act by increasing their positions to either change of VIX, which does not illustrate increased fear in their actions, but is rather evidence of noise trading behavior. Also, the findings of this study suggests that during the pandemic retail investors are more predictable, due to established market signals become “noisy” and those catch the attention of retail investors, such as VIX, or S&P500 returns as shown by Welch (2020).

Comparing with findings of Barber and Odean (2007) of retail investors attention, the findings of this thesis support that idea, by indicating that VIX during regular market conditions is not sufficiently noisy signal to attract retail investors’ limited attention. The limited attention of retail investors is related to herding as mentioned by Barber et. al (2020), and deducing from this, if a market indicator points to systematic behavior in the macrolevel, this could be seen as similar event as herding, but in a market wide

context. However, to make these conclusions and to confidently predict retail investor behavior, the indicators would need to provide more conclusive results.

Contrasting these results to established theories, we can see that during extreme market conditions, established market measures may contain information of retail investor activity, as they become noisy signals. To contrast with the weak form of efficiency (Malkiel, 1989), the results of this study suggest, that there could have been a possibility to predict a sliver of the retail investor activity based on VIX. Building the understanding upon these findings, retail investor behavior in the future might give arbitrageurs an opportunity to exploit these anomalies caused by retail investors.

6 CONCLUSIONS

6.1 Main Findings

Firstly, to conclude the main findings of this thesis, the answers to research question will be reviewed.

1. Does VIX provide predicting power over retail investor activity?

The results of the regressions show that during higher levels of VIX, the changes in direction offer different forecasting capabilities of retail investor activity. As during higher level of VIX, downturn increases the retail investor activity for following days which could be further evidence of retail investor providing liquidity to the market, as previous literature has shown evidence of trading volume correlating with VIX. Additionally, upswing causes a spike at the day of, and at five-day delay, which could be evidence of momentum seeking behavior. Though, the most robust conclusion based on the findings would be that with high enough level, VIX is a noisy signal that, whenever it experiences strong enough variance in either direction, retail investors act on it with a 5-day delay, which is further evidence of predictable systematic behavior of retail investors.

2. How does the level VIX affect retail investor activity?

The findings of this thesis points that, during extreme market conditions, VIX is a noisy signal that causes some variance in retail investor behavior. This relationship seems

to be non-existent during regular market conditions. It can be seen that during more extreme market conditions, and higher levels of VIX the retail investor activity has more economic value and can be better forecasted and measured. This adds to the evidence of previous literature, that retail investor activity has more economic significance during extreme conditions, and high levels of VIX usually occur at market turmoil.

3. How retail investors account for variation in VIX under different market conditions?

The findings suggest that retail investors activity is not related to variation in VIX under regular market conditions. However, during market turmoil retail investors experience VIX as a signal of noise as well as signal of fear, due to growth in in activity after a downturn in VIX for five consecutive days, where as a upswing only causes activity after five days.

To conclude, VIX includes information of retail investor activity and might provide additional explanatory power of retail investor activity during extreme market conditions, but as a singular variable, the variation of VIX only account for less than 30% of the variation in retail investor activity. These results provide further evidence of the increased economic value of retail investors during market turmoil and established market measures, such as VIX, may offer viable options for predicting retail investor activity.

6.2 Implications for International business

For variety of businesses understanding retail investor behavior may offer a way for excess returns, as foreseeing the behavior may offer opportunities to exploit the retail investor behavior. Understanding the retail investor activity under different market conditions can help with portfolio management, for seeing where retail investors pay attention to and how the stabilizing force may affect the risks. Additionally, understanding the market conditions and when and what retail investors pay attention to, may help timing issuance of stocks or a company going public. Financial institutions and businesses, who closely work with stock markets could improve on the retail investor predictions and develop models for other market conditions.

Additionally, the increased interest in investing and the lower barriers to investing may open opportunities for businesses to take advantage of this less experienced investor crowd.

6.3 Limitations

The research problem focused on recent market events and data from one brokerage firm. There is a lack of explanatory factors of retail investor activity, and this thesis wanted to shed light on one market explaining variable, and its relationship to retail investor activity. Other measures have been used in previous literature for models of predicting retail investor activity, and this thesis limits its research problem only to assess implied volatility as predictive variable. Other literature has shown statistical significance with other predictors, and the only one variable limits this research. The approach in this thesis is only quantitative, and to improve in further research, different variables could be combined to create better model, and develop a qualitative approach to improve the value of the information.

6.4 Suggestions for further research

This thesis was able to provide evidence of predictive power of VIX for retail investor behavior. The only methodology used was multilinear regression, and for which few assumptions were made, which could have affected the results. As mentioned in the previous section, to further improve the results, a qualitative approach and including a sample of retail investor activity with better representation would be beneficial. Properly testing the robustness of the regression model was not in scope of bachelor's thesis, but it would improve the value of the results. Additionally, combining the existing evidence from previous literature of variables with predicting power of retail investor behavior with the results of this thesis could be added to a more comprehensive model for forecasting retail investor behavior, such variables as price or return data of underlying assets or focus on specific industries to narrow down from macroeconomic perspective.

7. REFERENCES

Aharon, D.Y. and Qadan, M. (2020) When do retail investors pay attention to their trading platforms? *The North American Journal of Economics and Finance*, 53, p.101209

Albulescu, C. (2020) Coronavirus and Financial Volatility: 40 Days of Fasting and Fear. *SSRN Electronic Journal*

Andrikopoulos, P. (2005) Modern Finance vs. Behavioural Finance: An Overview of Key Concepts and Major Arguments. *SSRN Electronic Journal*

Ballinari, D. (2020) *Retail investors' trading activity and the predictability of stock return correlations*. p.

Barber, B.M., Huang, X., Odean, T. and Schwarz, C. (2020) Attention Induced Trading and Returns: Evidence from Robinhood Users. *SSRN Electronic Journal*

Barber, B.M. and Odean, T. (2001) The Internet and the Investor. *Journal of Economic Perspectives*, 15(1), pp.41–54

Barber, B.M. and Odean, T. (2007) All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors *Review of Financial Studies*, 21(2), pp.785–818

Barber, B.M. and Odean, T. (2013) The Behavior of Individual Investors. *Handbook of the Economics of Finance*, [online] 2, pp.1533–1570 Available from: https://faculty.haas.berkeley.edu/odean/Papers%20current%20versions/behavior%20of%20individual%20investors.pdf [Accessed 11 Feb. 2021]

Bates, A. (2020) *Investor Behavior in the Midst of a Global Pandemic Investor Behavior in the Midst of a Global Pandemic* [online] Available from: https://scholarworks.gvsu.edu/cgi/viewcontent.cgi?article=1794&context=honorsprojects [Accessed 11 Feb. 2021]

Boehmer, E., Jones, C.M. and Zhang, X. (2016) Tracking Retail Investor Activity *SSRN Electronic Journal*

Burghardt, M., Czink, M. and Riordan, R. (2008) Retail Investor Sentiment and the Stock Market *SSRN Electronic Journal*

Chiah, M. and Zhong, A. (2020) Trading from home: The impact of COVID-19 on trading volume around the world *Finance Research Letters*, 37, p.101784

Choi, H.M. (2019) Market uncertainty and trading volume around earnings announcements *Finance Research Letters*, 30, pp.14–22

Chorpenning, A. (2019) *Call vs Put Options: What's the Difference?* [online] finance.yahoo.com Available from: https://finance.yahoo.com/news/call-vs-put-options-difference-225058413.html

Christie, W. and Huang, R. (1995) Following the Pied Piper: Do Individual Returns Herd around the Market? *Financial Analysts Journal*, 51, pp.31–37

Desagre, C. and D'Hondt, C. (2020) Googlization and retail trading activity. *Journal of Behavioral and Experimental Finance*, p.100453

Dybvig, P.H. and Ross, S.A. (1989) Arbitrage. *Finance*, pp.57–71.

Economou, F., Philippas, N. and Caporale, G.-M. (2008) Herding behaviour in extreme market conditions: the case of the Athens Stock Exchange. *Economics Bulletin*, 7, pp.1–13

Ekkayokkaya, M., Jirajaroenying, S. and Wolff, C.C.P. (2020) Executing trades in style: retail investors vs. institutions. *Asia-Pacific Journal of Accounting & Economics*, pp.1–19

Fama, E.F. and French, K.R. (2004) The Capital Asset Pricing Model: Theory and Evidence. *Journal of Economic Perspectives*, 18(3), pp.25–46

Foucault, T., Sraer, D. and Thesmar, D.J. (2011) Individual Investors and Volatility. *The Journal of Finance*, 66(4), pp.1369–1406

Hirshleifer, D. (2001) Investor Psychology and Asset Pricing *The Journal of Finance*, [online] 56(4), pp.1533–1597. Available from: https://onlinelibrary.wiley.com/doi/10.1111/0022-1082.00379

Koo, B., Chae, J. and Kim, H. (2018) Does Internet Search Volume Predict Market Returns and Investors' Trading Behavior? *Journal of Behavioral Finance*, 20(3), pp.316–338

Lei, V.U.T., So, S.M.S. and Zou, M. (2012) Investor Sentiment - Relationship between VIX and Trading Volume *SSRN Electronic Journal*, 5(4)

Malkiel, B.G. (1989) Efficient Market Hypothesis *Finance* [online] pp.127–134 Available from: https://link.springer.com/chapter/10.1007/978-1-349-20213-3_13.

Moss, A., Naughton, J.P. and Wang, C. (2020) *The Irrelevance of ESG Disclosure to Retail Investors: Evidence from Robinhood* [online] papers.ssrn.com Available from: https://ssrn.com/abstract=3604847 [Accessed 1 Dec. 2020]

Nicolaus, D. (2010) *Derivative Choices of Retail Investors: Evidence from Germany* [online] citeseerx.ist.psu.edu Available from: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.225.2116&rep=rep1&type=pdf [Accessed 30 Nov. 2020]

Pagano, M.S., Sedunov, J. and Velthuis, R. (2020) *How Did Retail Investors Respond to the COVID-19 Pandemic? The Effect of Robinhood Brokerage Customers on Market Quality* [online] papers.ssrn.com Available from: https://ssrn.com/abstract=3703815 [Accessed 3 Jan. 2021]

Potter, V. and Whaley, R. (2008) *Understanding VIX* [online] Available from: http://www.growthpointinvestments.com/newsletters/images/UnderstandingVIX.pdf [Accessed 3 Jan. 2021]

Ramelli, S. and Wagner, A.F. (2020) Feverish Stock Price Reactions to COVID-19. *The Review of Corporate Finance Studies*

Ritual, A.M. (2020) *The Rise of the Online Investor During COVID-19 A Behavioural Finance Analysis into Retail Investor Sentiment*. [online] Available from: https://research-api.cbs.dk/ws/portalfiles/portal/62188126/881480_Final_version_of_the_Master_Thesis..pdf [Accessed 1 Dec. 2020]

Rooney, K. (2020) *Fintech app Robinhood is driving a retail trading renaissance during the stock market's wild ride* [online] CNBC Available from: https://www.cnbc.com/2020/06/17/robinhood-drives-retail-trading-renaissance-during-markets-wild-ride.html [Accessed 3 Jan. 2021]

Sarı, S. and Kartal, T. (2020) Covid19 Salgınının Altın Fiyatları, Petrol Fiyatları ve VIX Endeksi ile Arasındaki İlişki. , p

Schmittmann, J.M., Pirschel, J., Meyer, S. and Hackethal, A. (2014) The Impact of Weather on German Retail Investors* *Review of Finance*, 19(3), pp.1143–1183

Talwar, M., Talwar, S., Kaur, P., Tripathy, N. and Dhir, A. (2021) Has financial attitude impacted the trading activity of retail investors during the COVID-19 pandemic? *Journal of Retailing and Consumer Services*, 58, p.102341

Tripathi, G. (2014) An Empirical Investigation of Investors Perception towards Derivative Trading *Global Journal of Finance and Management*, [online] 6(2), pp.99–104 Available from: http://www.ripublication.com/gjfm-spl/gjfmv6n2_02.pdf [Accessed 30 Nov. 2020]

Welch, I. (2020) *Retail Raw: Wisdom of the Robinhood Crowd and the Covid Crisis* [online] www.nber.org Available from: http://www.nber.org/papers/w27866 [Accessed 1 Dec. 2020]

Whaley, R.E. (1993).Derivatives on Market Volatility. *The Journal of Derivatives*, 1(1), pp.71–84

Whaley, R.E. (2000) The investor fear gauge. *Journal of Portfolio Management*, [online] 26(3), pp.12–17. Available from: https://denvertradinggroup.com/wp-content/uploads/dtg-platinum/Trading/fear_gauge%20(VIX-VXN).pdf [Accessed on 5 Apr. 2021]

8. Appendices

Appendix 1: Screenshot of the RT data extract including first 20 tickers and first month of popularity data

timestamp	mtl_users_	joe_users_	dre_users_	amsf_user	lvhe_users	imh_users_	hauz_users	gfl_users_l	hdiv_users	clm_users_	nebu_user	otis_users_	sgb_users_	dcph_users	hall_users_
2.5.2018		55	205	42	13	41	6			1451			11	49	58
3.5.2018		58	206	42	13	40	6			1455			11	47	60
4.5.2018		58	206	42	13	40	6			1449			11	47	60
5.5.2018		55	205	42	13	39	6			1447			10	47	60
6.5.2018		55	205	42	13	39	6			1447			10	47	60
7.5.2018		55	207	42	13	39	6			1451			10	47	60
8.5.2018		56	210	43	13	39	6			1464			11	46	60
9.5.2018		55	211	43	14	40	6			1473			11	45	60
10.5.2018		55	212	43	13	41	6			1479			11	44	59
11.5.2018		54	212	43	13	41	6			1492			11	44	60
12.5.2018		54	211	43	13	41	6			1492			11	44	60
13.5.2018		54	211	43	13	41	6			1492			11	44	60
14.5.2018		56	225	43	13	41	6			1569			11	44	60
15.5.2018		55	225	42	13	41	6			1582			11	43	57
16.5.2018		55	222	42	13	41	6			1585			11	43	58
17.5.2018		55	221	42	14	42	6			1589			11	43	58
18.5.2018		55	228	42	14	40	6			1615			11	43	58
19.5.2018		54	225	42	13	39	6			1611			11	39	58
20.5.2018		54	225	42	13	39	6			1611			11	39	58
21.5.2018		54	225	42	13	39	6			1629			11	41	58
22.5.2018		54	226	42	13	39	6			1631			11	44	58
23.5.2018		53	228	41	13	39	6			1640			11	45	58
24.5.2018		54	228	40	13	39	6			1639			11	44	58
25.5.2018		55	226	40	14	39	6			1635			11	42	58
26.5.2018		55	222	40	13	39	6			1631			11	41	58
27.5.2018		55	222	40	13	39	6			1631			11	41	58
28.5.2018		55	222	40	13	39	6			1631			11	41	58
29.5.2018		55	222	40	13	39	6			1631			11	41	58
30.5.2018		54	215	40	15	38	7			1640			12	44	50
31.5.2018		55	215	40	15	38	7			1644			14	44	50
1.6.2018		56	216	39	15	38	7			1658			16	43	51
2.6.2018		56	216	39	15	37	7			1658			16	43	51

Appendix 2: Screenshot of the VIX data extract including first month of daily values.

Date	VIX Open	VIX High	VIX Low	VIX Close
2.5.2018	15,48	15,97	14,75	15,97
3.5.2018	15,78	18,66	15,43	15,9
4.5.2018	15,94	16,92	10,91	14,77
7.5.2018	15,32	15,52	14,51	14,75
8.5.2018	14,53	15,56	14,52	14,71
9.5.2018	14,54	14,63	13,38	13,42
10.5.2018	13,36	13,63	12,92	13,23
11.5.2018	13,22	13,44	12,65	12,65
14.5.2018	12,95	13,28	12,81	12,93
15.5.2018	13,13	15,01	12,5	14,63
16.5.2018	14,38	14,91	13,21	13,42
17.5.2018	13,54	13,86	12,65	13,43
18.5.2018	13,18	13,87	13,06	13,42
21.5.2018	13,44	13,59	12,78	13,08
22.5.2018	13,03	13,42	12,77	13,22
23.5.2018	13,5	14,6	12,49	12,58
24.5.2018	12,73	14,24	12,53	12,53
25.5.2018	12,44	13,52	12,29	13,22
29.5.2018	14,39	18,78	14,39	17,02
30.5.2018	16,6	16,64	14,65	14,94
31.5.2018	14,93	16,29	14,2	15,43
1.6.2018	14,92	14,93	13,37	13,46

Appendix 3: Screenshot of the first month's values of the variables of the pandemic timeframe.

Date	Rtchange	VIX-	lag1VIX-	lag2VIX-	lag3VIX-	lag4VIX-	lag5VIX-	VIX+	lag1VIX+	lag2VIX+	lag3VIX+	lag4VIX+	lag5VIX+
2.1.2020	0,65 %	7,64 %	7,41 %	0,00 %	0,00 %	0,71 %	0,00 %	0,00 %	0,00 %	7,57 %	6,30 %	0,00 %	0,16 %
3.1.2020	0,42 %	6,82 %	7,64 %	7,41 %	0,00 %	0,00 %	0,71 %	0,00 %	0,00 %	0,00 %	7,57 %	6,30 %	0,00 %
6.1.2020	0,00 %	10,93 %	6,82 %	7,64 %	7,41 %	0,00 %	0,00 %	0,00 %	0,00 %	0,00 %	0,00 %	7,57 %	6,30 %
16.1.2020	4,39 %	0,00 %	10,93 %	6,82 %	7,64 %	7,41 %	0,00 %	0,98 %	0,00 %	0,00 %	0,00 %	0,00 %	7,57 %
17.1.2020	0,40 %	0,90 %	0,00 %	10,93 %	6,82 %	7,64 %	7,41 %	0,00 %	0,98 %	0,00 %	0,00 %	0,00 %	0,00 %
21.1.2020	0,97 %	2,91 %	0,90 %	0,00 %	10,93 %	6,82 %	7,64 %	0,00 %	0,00 %	0,98 %	0,00 %	0,00 %	0,00 %
22.1.2020	0,31 %	0,00 %	2,91 %	0,90 %	0,00 %	10,93 %	6,82 %	3,63 %	0,00 %	0,00 %	0,98 %	0,00 %	0,00 %
23.1.2020	0,22 %	2,13 %	0,00 %	2,91 %	0,90 %	0,00 %	10,93 %	0,00 %	3,63 %	0,00 %	0,00 %	0,98 %	0,00 %
24.1.2020	0,25 %	0,00 %	2,13 %	0,00 %	2,91 %	0,90 %	0,00 %	13,27 %	0,00 %	3,63 %	0,00 %	0,00 %	0,98 %
27.1.2020	0,62 %	0,00 %	0,00 %	2,13 %	0,00 %	2,91 %	0,90 %	4,54 %	13,27 %	0,00 %	3,63 %	0,00 %	0,00 %
28.1.2020	0,31 %	3,97 %	0,00 %	0,00 %	2,13 %	0,00 %	2,91 %	0,00 %	4,54 %	13,27 %	0,00 %	3,63 %	0,00 %
29.1.2020	0,41 %	0,00 %	3,97 %	0,00 %	0,00 %	2,13 %	0,00 %	4,43 %	0,00 %	4,54 %	13,27 %	0,00 %	3,63 %
30.1.2020	0,23 %	14,01 %	0,00 %	3,97 %	0,00 %	0,00 %	2,13 %	0,00 %	4,43 %	0,00 %	4,54 %	13,27 %	0,00 %
31.1.2020	0,21 %	0,00 %	14,01 %	0,00 %	3,97 %	0,00 %	0,00 %	14,79 %	0,00 %	4,43 %	0,00 %	4,54 %	13,27 %