POLITECNICO DI TORINO

MASTER's Degree in MANAGEMENT ENGINEERING

MASTER's Degree Thesis

The rise of online trading platforms during highly volatile markets: retail investors and a comparison with institutional investors

Supervisor

Prof. Riccardo CALCAGNO

Candidate

Avan FATAH

July 2021

Summary

Online trading platforms enabled a new way of investing. Now, placing an order on the financial markets only requires a device and an internet connection, there are no particular requirements and even low amounts of money are enough to make a transaction, furthermore, platforms are designed for the public and therefore have a user-friendly interface and low or zero fees. Those characteristics make the platforms attractive and accessible for small investors with any financial background. Retail investors might engage in investing activities without being aware of the risks and regulators' attention should be required.

The following study aims at analysing this emerging group of low informed investors by comparing them with the rational investors: the institutions. The analysis is based on two set of stocks: the most popular stocks among the retail investors and the most popular among the institutions, for the retail investors the analysis is performed on two time horizons: one that includes the high volatility time of the Covid outbreak, and one that excludes it and represents a normal condition market.

Both groups' investment activity is evaluated in relation with four market conditions: S&P500, VIX, stock price and market volumes (S&P500 is

excluded in the regression due to the presence of multicollinearity). A preliminary assessment is made from the Spearman correlation, then a deeper assesment is made with the regression for panel data and the multiple linear regression. The results show that retail investors invest during high market volatility, while the institutions are risk averse, the result can be explained by the demographic characteristics and the psychological biases that affect individuals: the retail investors of the study, which come from the Robinhood Market Inc. (RH) platform database, are young, have a low account size and might be influenced by social media and low quality information sources, their decisions are likely affected by the level of attention (raised by news, social media, announcements), the sunk cost trap, the disposition effect and the aversion to losses. Institutional investors instead have dedicated time, resources and a defined strategy, therefore they unlikely fall in those traps (Barber and Odean 2008).

For both, the investment activities are positively related to market volumes, meaning that they invest when there is market movement, and regarding prices, retail investors tend to follow a contrarian and a momentum strategy, the first is followed for stocks which have low financial performances.

Regarding the portfolio, the composition and performance is compared to the optimal portfolios on the efficient frontier, first for the RH investors holding the most popular stocks and then for RH investors holding the institutional stocks. With respect to the institutional, the retail portfolio has higher Sharpe ratio and higher volatility due to the intrinsic stock's characteristics, however in both the portfolio types, RH investors hold an excessive amount of few volatile stocks that increase their portfolio risk without contributing to the returns. Overall, the most surprising fact is that unsophisticated investors get more involved in financial activities when the market is volatile.

Acknowledgements

Thanks to my supervisor, Prof. Riccardo Calcagno, who followed me during the thesis work, and to the Professors encountered during my studies, who provided a high quality teaching and raised my interest and curiosity for their subjects. Thanks to my friends, who were always supportive.

Table of Contents

List of Tables

List of Figures

Acronyms

RH

Robinhood Markets Inc.

YF

Yahoo Finance

FTH

Full-time horizon

PCH

Pre-Covid time horizon

BG

institutions with highest AUM

BK

Bank

HF

Hedge fund

IN

Insurance

PF

Pension fund

Chapter 1

Literature review

The relation between volatility and investors is found extensively in literature, Foucault et al. (2011) find that retail investors behave as noise traders (or liquidity traders) and are a determinant of volatility, with Barber et al. (2008) they support the fact that retail investors have poor performance. De Long et al. (1990) show that irrational investors increase and bear the market risk they create, enabling them instead, to earn more than rational investors, and contributing to driving the price away from the fundamental values. Regarding rational investors, Gabaix et al. (2007) find that large institutions can generate large movements in the markets and can be a cause of volatility (this effect can be mitigated if institutions split their orders).

Those studies consider investors' behaviour as one of the determinants of market volatility. However, during the Covid outbreak, retail investors seemed to be 'attracted' by volatile markets, with participation rapidly increasing during this high volatility time. Recent studies relate investors, market participation and volatility through attention.

Audrino et al. (2020) find that attention and sentiment variables, measured by Google searches and StockTwits messages, have explanatory power on the volatility and can indeed be used to improve volatility prediction models, also Smales (2021) finds that higher level of attention leads to a quicker information incorporation in markets and is associated with volatility. Ballinari et al. (2019) include also institutional investors, both retail and institutional investors' attention can impact market prices, while for the first, attention is positively related to post-announcement volatility and to a slower price adjustment, for the second, attention is slightly negatively related to post-announcement volatility, and leads to a quicker price adjustment. Retail investors are indeed likely to misinterpret the information and generate more disagreement. Aharon and Qadan (2020), based on a dataset between 2014 and 2017 (therefore it does not include Covid outbreak), focus on traders' attention to their trading platform, and find that market shocks (i.e. highly volatile markets) increase retail investors' attention to their trading platform and to the financial market and lead them to get more involved in information gathering.

Previous research estimate retail investors' attention, sentiment and behaviour from measures such as surveys or more recently, Google searches, Google trends, news' visits or social media messages. This study is based on the novel database that provides a direct measure of retail investors' behaviour, that is users movements on the online trading platform Robinhood (abbreviated as RH), who can be profiled with additional details (type of trading platform, financial background, age, account size, country, information sources).

Institutional data comparable to users' number come from the SEC's 13F filings. Another characteristic is the considered time-horizon, which includes a period of very high volatility, the Covid outbreak.

Few and very recent papers used this database and addressed Covid outbreak, Beck and Jaunin (2021) find that retail investors can provide liquidity during crisis times and have an impact on the market prices, this effect is amplified due to institutional investors' price inelasticity, also Welch (2020) finds that retail investors had stabilizing role during market volatility, since they did not show panic and that they had a good portfolio performance, Eaton et al. (2021) find that while retail investors on the aggregate have a positive impact on the market quality, RH investors behave as noise traders and increase market volatility and also for Baig et al. (2021) they contributed to destabilizing the market. The studies have the aim of evaluating the impact of retail investors on the financial markets, instead, the following work aims at analysing how investors, retail and institutional groups, are related to each other and how they react to market conditions, distinguishing normal and volatile markets and comparing them at the level of the market and their portfolio.

Regarding the psychological biases that can influence investors, Daniel Kahneman, the 2002 Nobel Prize in Economic Sciences, gave an important contribution to behavioural economics. In his book 'Thinking, fast and slow' (2011) he summarizes his research: the human brain is the outcome of a long evolutionary process, where natural selection determined the 'successful' human characteristics that allowed survival. The human brain is still subject to responses which were essential in the past, and today are still important but can be misleading. Those are the intuitive, immediate reactions (of the brain's 'system 1') to events which lead to take decisions without involving a deeper reasoning (the 'system 2'), and not always people are conscious about this process.

The prospect theory explains mathematically how individual decision making violates the rational principles, a decision is indeed influenced by how it is formulated and by the aversion to losses, the utility as a function of gains and losses shows a diminishing sensitivity (if individuals already beared a high gain or loss, the utility coming from a marginal increase or decrease in losses or gains is lower) and a convex curve in the losses quadrant (individuals overweight losses). This explains many psychological biases which affect people in any decision making, also in the financial field.

Institutional investors have defined tools, strategies and expertise, they rely on automatization systems and therefore are unlikely to be subject to irrational decisions and biases, however, the opposite holds for retail investors, who as human beings, are adverse to losses and in the financial markets want to have instant gratification and secure their profits (Barber and Odean, 2008), therefore, to make early profit they may sell good stocks which will likely perform better in the future, and keep bad stocks, to avoid to realise the loss ('disposition effect'), and probably believing it will go better in the future (overconfidence), while instead, without selling unprofitable stocks, they miss the opportunity of using that money for making a better investment. (Odean, 1998; Shefrin and Statman, 1985). Another reason for which investors tend to keep bad stocks is the sunk cost bias, that leads to continue to pursue the same strategy only because an amount of effort

was already spent, while instead sunk cost should not be considered in the decision making (Blumer,1985).
Other biases are related to optimism and overconfidence that lead to excessive trading (Daniel and Hirshleifer, 2015) or gambling and preference for state lottery-like stocks, which have low price, are risky and have low expected returns and attracts retail investors which are young, with low income and live in regions with high unemployment (Kumar, 2009), limited attention and investment decision based on familiarity can lead to hold concentrated and local portfolios, which are not optimal (Korniotis and Kumar, 2008).

Chapter 2

Data

2.1 Robinhood Markets Inc

Robinhood Markets Inc (abbreviated as RH), is an online trading platform pioneer. It was founded in 2013 with the objective of allowing an easy and cheap way to participate in financial markets.
Indeed, it offers commission-free trading and users can start to invest even with low amount of money, the platform is accessible through its mobile application or desktop website, it has a user-friendly interface, and it provides real-time information about stocks such as ratios, prices, volumes, news, popularity (which is a real-time list of most widely held stocks among the users). The tradable instruments are stocks, ETFs, ADRs, options, and cryptocurrencies. It currently accepts only US subscriptions.
The company makes revenues mainly from payments for order flow, in which the broker receives a compensation for directing orders to different parties for trade execution. Other sources of income are the 5$ monthly fee for optional

membership to Robinhood Gold (which gives the client access to additional investing tools and information), interest on uninvested cash; lending stocks purchased on margin; and fees on purchases using the company's debit card. Recently the company was subject to several lawsuits, in December 2020, it solved a civil fraud investigation by paying $65 million, initiated by the Securities and Exchange Commission (SEC) due to its missing disclosing of its practice of payment for order flow. Again in December, Massachusetts securities regulators filed a complaint blaming RH for exposing investors to needless risks associated with trading, by the aggressive market practice toward inexperienced investors and failing to ensure that proper protection controls were implemented[1].

As the data below shows, RH is relatively small with respect to its competitors: it has lower assets under management, number of users and account size of users. In the last years it faced a rapid increase of users, transactions' number and consequently, also revenues [2].

Launch date	18 April 2013
Headquarter	Menlo Park, California
People	Vladimir Tenev and Baiju Bhatt (co-CEO's), Jason Warnick (CFO), Gretchen Howard (COO)
Industry	Stockbroker and cryptocurrency
Company type	Private

Table 2.1: Robinhood Market Inc. company. Source: Businessofapps.

[1]https://www.investopedia.com/articles/active-trading/020515/how-robinhood-makes-money.asp

[2]https://www.businessofapps.com/data/robinhood-statistics/

Year	Total transactions	Company Valuation	Total Users
2018	$100 billion	$7 billion	6 million
2019	$150 billion	$11.7 billion	10 million
2020	$350 billion	$20 billion	20 million

Table 2.2: Robinhood Market Inc. growth. Source: Businessofapps.

	assets under management*	average account size
Robinhood	$20 billion	$3,500
E-trade	$600 billion	$100,000
TD Ameritrade	$1.3 trillion	$110,000
Charles Schwab	$3.8 trillion	$240,000

* Assets under management means the total market value of all investments a company holds.

Table 2.3: Robinhood Market Inc. and competitors. Source: Businessofapps.

2.1.1 Robinhood investors' profile

With data coming from web traffic trackers, specifically Alexa and SimilarWeb, it is possible to get insights on the behavioural traits of investors. The websites visited by users before or after visiting RH platform are social networks (Facebook, Youtube, Twitter, Reddit) and financial information websites (Yahoo Finance, Investopedia, Coinbase). This is indicative for the sources of information leading RH investors decision-making about their investments. Channels like Youtube have several content related to trading, from basic tutorials about platforms and financial markets to investment strategies tips. Social networks like Facebook or Reddit have an influence in driving investors' attention toward certain stocks and influence RH decisions, not necessarily in a rational way. Investopedia provides financial contents, mainly directed to financial education, this, together with the most popular FAQs visited (frequently asked question), can support the fact that RH

investors do not have a strong financial background.
Users' account size is relatively small, it ranges from 1000$ to 5000$, therefore they may hold a concentrated portfolio with a limited amount of stocks, or instead, seek for low price per share stocks and buy a limited amount of each. Their average age is 31, and they are mainly US based (indeed RH platform only accepts US subscriptions), and spend around 11 minutes on the platform.
It is reasonable to consider RH users to represent retail investors' behaviour. Retail investors are non-professional investors who invest their own money for themselves, they are usually driven by personal goals and have a small purchasing power. Not necessarily they are unsophisticated investors, for example among them there can be informed investors who have a financial background, however, given the information above, it can additionally be said that RH investors are retail investors without a solid financial knowledge, they are young and likely to be risk-prone (they might see trading and investment activity as a secondary source of revenue, a 'game' or a new experiment), and use as sources of information social networks and financial educational websites.

Rank	Robinhood FAQ Category	Robinhood Visits /1,000	Other Retail Brokers FAQ Category	Other Retail Brokers Visits /1,000
1	What is the Stock Market	6.49	What are Stock Splits	1.67
2	What is the DJIA	6.07	What is an ETF	1.48
3	What is the S&P 500	5.78	What are Puts and Calls	1.45
4	What is a PE Ratio	5.73	What are the Different Order Types	1.41
5	What are Different Order Types	4.96	How to Trade IPOs	1.32
6	What is a Fiscal Year	4.72	What is RSI	1.25
7	What are Extended Hours	4.36	How to Find Investments	1.22
8	How to Trade / Invest	4.24	How are Investments Taxed	1.20
9	How to Find Investments	3.97	Mutual Funds vs ETFs	1.15
10	What is Pattern Day Trading	3.83	Trading Fees	1.14

Table 2.4: Robinhood Market Inc. most visited FAQs (Beck and Jaunin 2021)

2.2 Securities and Exchange Commission (SEC)

SEC is a US independent federal government entity which supervises the stock exchange. It was found by the Congress in 1934 to restore market confidence in response to the stock market crash of 1929 that led to the Great Depression, and its first chairman was Joseph P. Kennedy. SEC's main functions are of monitoring and promoting fairness in the securities markets, the aim is to protect investors over unlawful market actions and facilitate the flow of information on companies and professionals to help investors make informed decisions. Entities in the securities markets (including securities exchanges, brokerage firms, dealers, investment advisors, and investment funds) must comply with SEC regulation and provide registration statements, financial reports and securities forms which are accessible to the public through SEC's online database, EDGAR (Electronic Data Gathering, Analysis, and Retrieval)[3] [4].

13F filings:

Among the available forms, the 13F filings were used to develop the analysis about institutional investors. They are filed quarterly, and contain the institutional holdings (shares, options, notes, bonds) fair value and volume at each quarter. However not all holdings are required to be reported, institutions must report only long positions holdings, which have an aggregate fair market value of at least 100,000,000$, within 45 days after the last day of each of the first three calendar quarters[5].

2.3 Data collection

Data were collected, merged and processed through the programming language Python. For retail investors, they were gathered from Robintrack, made available by the author Casey Primozic by connecting to the RH API. The dataset includes 8597 csv format files, each corresponding to a stock and containing the number of users who were holding at least one share of the company, with an almost hourly frequency, for a time horizon of about 2 years: from 02 May 2018 to 13 August 2020, when the RH API was closed. Daily data about stocks and market (stock price, market of reference volumes, VIX, S&P500) were collected from Yahoo Finance API (YF).

[3] https://www.sec.gov/about/what-we-do
[4] https://www.investopedia.com/terms/s/sec.asp
[5] https://www.sec.gov/pdf/form13f.pdf

Given the different and, for RH irregular, frequencies, first the RH data were resampled daily by mean and then merged with the YF data. From the two datasets merge, it was possible to create a database containing daily information for each stock, which enables to analyse how users behave respect to market conditions (for each stock, daily number of users, price, market volume, VIX, S&P500), the dataset was cleaned removing all closing days (weekends, holidays) and removing null values, the total number of days in each dataset is 584.

Data about institutional investors were collected from the 13F filings in the SEC database EDGAR. SEC data are not in form of downloadable file, each institution and each quarter have to be accessed one by one, and each saved to a file, therefore, the data collection procedure was far more complex with respect to RH.

Each filing is referred to with a reporting date and a filing date, the reporting date was taken as a reference for the analysis since it represents the 'real time' holding, while the filing date corresponds to the date in which data were disclosed, between 45 days to 5 months after the investment took place. Each quarter contains the companies' holdings names, the number of holdings, their class (shares, options, notes, bonds), and the fair market value. A single holding is divided in many splits, therefore, firstly, the data was filtered in order to only consider shares, secondly it was grouped by stock to get the overall amount of shares for each stock in each quarter. The files were aggregated together to have a dataset containing all the quarters' data for each institution. However, the fair market value is a subjective value, usually the non-adjusted market price, which does not give an indication about the acquisition price. Therefore, data were aggregated with YF (resampled quarterly) in order to create a dataset with quarterly information about stock holdings, market price and market conditions, in the two years horizon for each institution of interest.

Chapter 3

Robinhood users

The analysis on RH users considers two scenarios:

1. The full-time horizon (abbreviated as FTH), which contains the whole available data and includes the period of highly volatile markets, Covid outbreak. It covers the time from 08/05/2018 to 13/08/2020.

2. The Pre-Covid horizon (abbreviated as PCH), which does not include the Covid outbreak and represents a period of normal condition markets, without any significant volatility change. It covers the time from 08/05/18 to 23/02/20

The beginning of the Covid outbreak is considered to be the 24/02/20, the day in which the VIX started to increase and when the Dow Jones Industrial Average and FTSE 100 dropped more than 3% as the Covid outbreak worsened in globally[1]

[1]https://www.bbc.com/news/business-51612520

The following analysis is based on the 20 most popular stocks among the RH users. First, a preliminary assessment on the RH users behaviour is made through the graphical representation and the Spearman correlations between the users and the market factors.
Then, a deeper analysis is made with the regressions: the analysis for panel data is performed to evaluate 'globally' how users behave with respect to the market conditions, then, the multiple linear regression is performed to evaluate more specifically the behaviour for each stock.
The last step evaluates the RH portfolio performance and compares it with the optimal portfolio on the efficient frontier.

The market factors are S&P500, VIX, prices and volumes:
S&P500 is a market index reflecting the performance of the 500 largest companies that are listed on the US stock market, VIX (Cboe Volatility Index) indicates the market expectations regarding the volatility, it is based on the prices of SPX index options and represents market sentiment, the volumes are the stock's volumes of the overall market of reference for each stock (mainly Nasdaq and NYSE), and the prices are the market prices of each stock.

Those measures were chosen to evaluate a relation between RH users, used as a proxy for retail investors, and the market conditions, firstly because they represent the main market factors: market performance, volatility, movement and price paid for the stock, and allow to evaluate if institutional and retail investors invest when the market is going well or bad, is stable or risky, when there is high or low interest in a stock, when the stock price is increasing or decreasing; secondly, they are publicly available and easy to find and understand without particular searching efforts or financial background, therefore investors may actually refer to those metrics for their investment decisions. However, as seen in the analysis, investors are related to those indicators but there is no evidence of a causal relationship that indicates they follow the metrics, or that the market condition causes their behavior.

3.1 Stock popularity

In order to find a rank representing the quarterly stocks' popularity on the platform, RH users' data were resampled quarterly, then aggregated to a matrix and transposed in order to have for each date, the stocks' name and users' number. By sorting in descending order the users in each date, it was possible to get a rank representing stock popularity. During the time horizon of 2018-2020, the first 10 most popular stocks on RH platform in each quarter were the following 20 stocks (the detailed number of users and classification per each quarter is found in the Table A.1):

Company name	Ticker symbol	Country ISO code	Sector
AMAZON.COM, INC.	AMZN	US	Retail
APPLE INC.	AAPL	US	Computer Hardware
MICROSOFT CORPORATION	MSFT	US	Industrial, Electric & Electronic Machinery
FORD MOTOR CO	F	US	Transport Manufacturing
FACEBOOK, INC.	FB	US	Media & Broadcasting
GENERAL ELECTRIC COMPANY	GE	US	Transport Manufacturing
WALT DISNEY COMPANY (THE)	DIS	US	Media & Broadcasting
TESLA, INC.	TSLA	US	Transport Manufacturing
NETFLIX, INC.	NFLX	US	Travel, Personal & Leisure
AMERICAN AIRLINES GROUP INC.	AAL	US	Transport, Freight & Storage
DELTA AIR LINES, INC.	DAL	US	Transport, Freight & Storage
ADVANCED MICRO DEVICES INC	AMD	US	Industrial, Electric & Electronic Machinery
CARNIVAL CORPORATION	CCL	PA	Transport, Freight & Storage
TWITTER, INC.	TWTR	US	Media & Broadcasting
SNAP INC.	SNAP	US	Industrial, Electric & Electronic Machinery
FITBIT, INC.	FIT	US	Industrial, Electric & Electronic Machinery
GOPRO, INC.	GPRO	US	Industrial, Electric & Electronic Machinery
AURORA CANNABIS INC.	ACB	CA	Chemicals, Petroleum, Rubber & Plastic
CRONOS GROUP INC.	CRON	CA	Chemicals, Petroleum, Rubber & Plastic
PLUG POWER INC	PLUG	US	Industrial, Electric & Electronic Machinery

Table 3.1: Top10 stock popularity on RH platform. Sector source: Orbis.

The following figure shows a plot over the entire horizon of the number of RH users holding at least one stock for each of the popular companies. From a preliminary qualitative evaluation, it is shown that RH users invest mainly in stocks which are well known by the general public. Surprisingly, the rapid increase of users starts around the Covid19 outbreak, a time in which markets were crashing and had high volatility, and in which small investors should not participate, given the high risks. The following analysis will investigate about the relation among users and market conditions, including the volatility.

Other patterns that can be observed are the a sudden increase in popularity, starting right at the beginning of the Covid outbreak, of stocks that were likely to face a temporary price decrease, such as AAL, DAL, CCL in the travel sectors, those stocks did not have many users until February 2020, and RH users, probably driven by overconfidence and optimism, heavily invested in those stocks despite the risks of bankrupcy that companies could face,

indeed, the outbreak mostly impacted on the travel and turism sector since people were prevented to travel, and not necessarily they could recorver after this business impact.
Users are investing also in PLUG, it is likely that they are following the hydrogen trend in the automotive field, which is a promising emergent technology but there is uncertainty about its future evolutions. Regarding the users' trends, AAPL, MSFT, TSLA, AMZN show a continuous increase of users for the whole time horizon, without any hint of decrease even at the end of the horizon, users might be confident that the prices of those stocks will continue to increase.
F and GE show a rapid increase during the Covid outbreak, but differently from the previous stocks, they show a decrease at the end of period, and AAL, DIS, GPRO, PLUG, SNAP, CCL, NFLX, between April 2020/May 2020 until the end of the time horizon face a decrease too, instead, FB, TWTR,CRON,AMD have not much movement.

Many of the popular stocks are being sold at the end of the horizon, consistently with the literature findings (Shefrin and Statman 1985; Odean 1998) for which retail investors need an instant gratification and to secure profits and tend to sell too soon their profitable stocks.

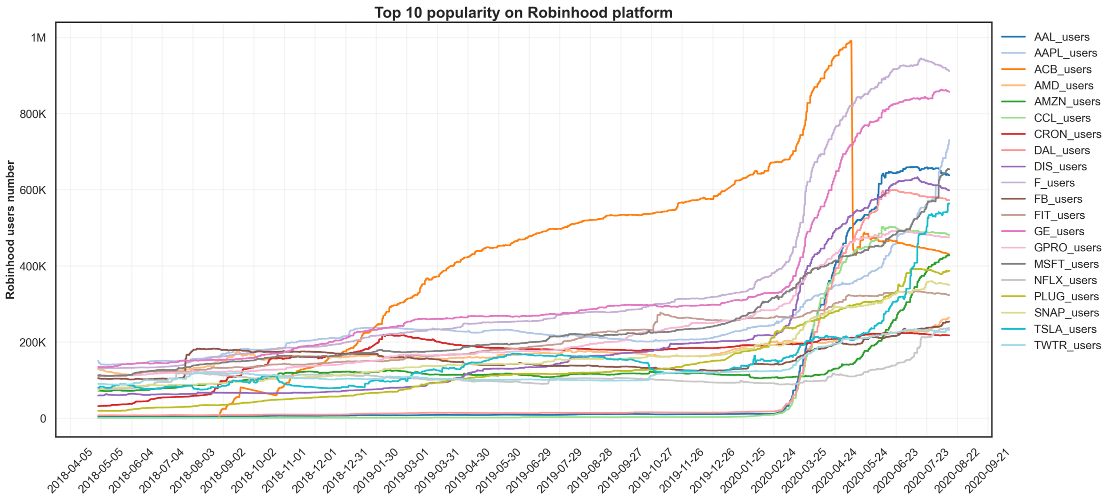

Figure 3.1: Top10 stock popularity on RH

3.2 Correlation

RH data is not normally distributed, therefore the correlations between RH users and the other factors are calculated with the natural logarithm transformation.

The correlation type used is the Spearman rank correlation, which does not require assumptions on the distribution, it attributes a rank to the data and detects any monotonic relationship. As expected, the Spearman correlation among RH users is strongly positive, except for ACB stock, indicating that overall users behave in the same way for the popular stocks.

The correlation is then performed for each stock, both in the PCH and the FTH, between RH users and the market conditions S&P500, VIX, market volumes and market prices. In both time horizons the correlation with prices is strong, suggesting that RH users follow prices. A relevant difference is on the VIX: while before Covid the correlations were weak and for some stocks also negative, in the FTH the correlation is stronger and positive for almost all the stocks (Table A.2).

Also, there is a positive relation with market volume (stronger in the FTH), consistent with the literature findings for which investors invest when attention is high (Aharon and Qadan 2020), and when attention is high volumes are also high (Alanyali et al. 2013), high volumes only indicate market movement, in any case, whether market is selling or buying the stock, investors participation increases.

3.3 Regression

Panel data, also called longitudinal data, are data models in which there are multiple entities and each entity is observed at multiple time periods.
Among the possible models, pooled, random effects and fixed effects regression, the latter is more suitable for the RH data. It is an extension of the multiple regression that exploits panel data to control for variables that differ across entities but are constant over time. From the notation $Y_{it} = \beta_1 X_{1it} + ... + \beta_k X_{kit} + \alpha_i + u_{it}$ where $i = 1, ..., n; t = 1, .., T$ The assumption to hold are the following:

1. u_{it} has conditional mean zero: $E(u_{it}|X_{i1}, ..., X_{iT}, u_{i1}, ...u_{iT}, \alpha_i) = 0$.

2. $(X_{i1}, ..., X_{iT}, u_{i1}, ...u_{iT}), i = 1, .., n$ are independent and identically distributed (i.i.d.) draws from their joint distribution.

3. Large outliers are unlikely: (X_{it}, u_{it}) have nonzero finite fourth moments.

4. There is no perfect multicollinearity.

For multiple regressors, the model allows the independent variables to be autocorrelated within but not among each other.

The notation is the following, where β_0 is the common intercept, $users\_pct_{it}$ the dependent variable percentage change of RH users for stock i at time t, and the the independent variables corresponding to $price_{it}$, VIX_t and $volume_{it}$, with $\gamma_j Dj_i$ being the unobserved variable that captures factors which are individual to the stock, $D2_i = 1$ if $i = 1$ and is 0 otherwise and so forth, u_{it} is the error term. $S\&P500$ was excluded given the positive correlation with prices.

$$users\_pct_{it} = \beta_0 + \beta_1 VIX_t + \beta_2 price_{it} + \beta_3 volume_{it} + \gamma_2 D2_i + \gamma_3 D3_i + u_{it}$$

In the FTH, at the 5% confidence level, the independent variable coefficients are significant and positive for all the variables price, VIX, and volume. In PCH, only VIX and volume are significant, and not the price, differently from the FCH, the coefficient of VIX is negative, and R^2 is lower.

This suggests that in 'normal' markets, where there are no significant events or high volatility episodes, the retail investors are less concerned with the market conditions VIX, volumes and stock prices, and tend to be risk averse. Instead, in times of high volatility, they are more concerned with market conditions and, surprisingly, invest more if the market volatility increases.

full-time horizon					
	coefficient	T-statistic	P-value	R-squared	F-statistic
users_pct				0.054	0.000
const	-0.102	-19.2	0.000		
price_ln	0.002	3.861	0.0001		
VIX_ln	0.006	11.567	0.000		
volume_ln	0.005	16.514	0.000		

pre-Covid horizon					
	coefficient	T-statistic	P-value	R-squared	F-statistic
users_pct				0.012	0.000
const	-0.054	-8.03	0.000		
price_ln	0.001	1.317	0.188		
VIX_ln	-0.002	-2.487	0.013		
volume_ln	0.004	10.139	0.000		

Table 3.2: RH users regression for panel data

Further analysis can be made on the single stocks to capture specific behaviours. The multiple linear regression aims at explaining a dependent variable Y with multiple independent variables X_i.

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

The least square assumptions are the following:

1. u_i has conditional mean zero: $E(u_i|X_i = 0)$.

2. $(X_i, Y_i), i = 1, .., n$ are i.i.d. draws from their joint distribution.

3. Large outliers are unlikely: (X_i, Y_i) have nonzero finite fourth moments.

The regression is performed for each of the 20 stocks in both the time horizons (the table Table A.3 shows a summary of all the regressions outputs). In the notation, $users\_pct_t$ are the percentage change of users in time t, and the independent variables are VIX, $price$ and $volume$ at time t with u being the error term.

$$users\_pct_{it} = \beta_0 + \beta_1 VIX_t + \beta_2 price_{it} + \beta_3 volume_{it} + u_{it}$$

Time series present the problem of non-stationarity and autocorrelation. The OLS linear regression tests show that some assumptions do not hold: the Durbin-Watson test indicates that there is positive autocorrelation among the residuals (values between 0 and 2) and the Jarque-Bera test indicates that the residuals are not normally distributed (p-value lower than 0.05).

The transformation used was percentage change for the dependent variable RH users, which eliminates the linear increasing trend and the seasonality, while the independent variables were transformed with the natural logarithm to reduce their variability. Using percentage change, logarithm transformation or differentiation for both the independent and the dependent variables, the output does not get better and for both test the same results are obtained.

The findings confirm the previous analysis: with respect to the FTH, PCH has a mainly negative relation with VIX, and has a lower significance for the variables, with lower coefficients and R-squared (average R-squared of 0.13 in PCH and 0.23 in FTH). In both horizons the relation with volume is positive and significant.

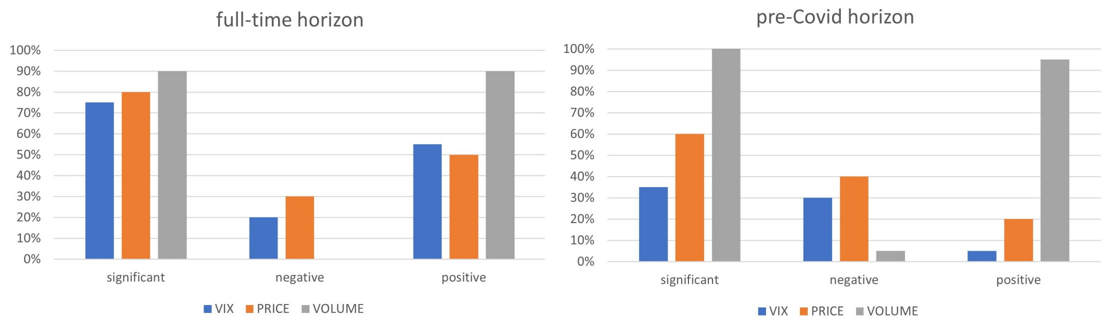

Figure 3.2: RH multiple linear regression

Also the linear regression models do not have much explanatory power, and do not show a causal effect. Nevertheless, the results can be useful to derive the direction of the relation and compare different time horizons and stocks. However, a possible reason for the low explanation might be that users tend to invest during high volatility not because of the market condition, but because of something else that happens together with those market conditions, for example news attracting their attention. Otherwise, it would be expected that also before the Covid pandemic there were a positive coefficients with VIX and a higher explanatory power.
The attention levels can drive retail investors financial behaviour: as shown by Alanyali et al. 2013 in highly volatile or high volumes markets, news and media are more focused on markets leading retail investors to pay attention to their trading activities or also, attract non-investors individuals in the financial markets. Indeed, as the 'total users' in Table 2.2 shows, the increased number of users buying a specific stock is not coming only from existing users, but also from new users that joined the platform for the first time.

3.4 Portfolio

Considering the prices from the multiple linear regression, RH investors pursue both a contrarian and a momentum strategy, in the first case, when users are negatively related to prices, the tendency is to buy bad performing stocks and sell well performing ones (Jegadeesh and Titman 1993), in the second case, when users are positively related with prices, the tendecy is to buy the well performing and sell the bad performing ones (De Bondt and Thaler 1985).
RH users pursue a contrarian strategy in the category Industrial, Electric & Electronic Machinery (FIT, GPRO, PLUG, SNAP) with CRON belonging to Chemicals, Petroleum, Rubber & Plastic and TWTR to Media & Broadcasting, for the other categories the strategy is momentum.
By evaluating the stock's financial performance over the years 2018-2020 (Figure A.1), it is found that the contrarian strategy is followed for 'non promising' stocks, which have not very attractive financials: they have indeed low levels of revenues, cash flow, total assets, ROE and ROA, also, they have a high level of liquidity relatively to the considered stocks, which might not be beneficial since a too high level indicates that the company is too liquid and is not investing.

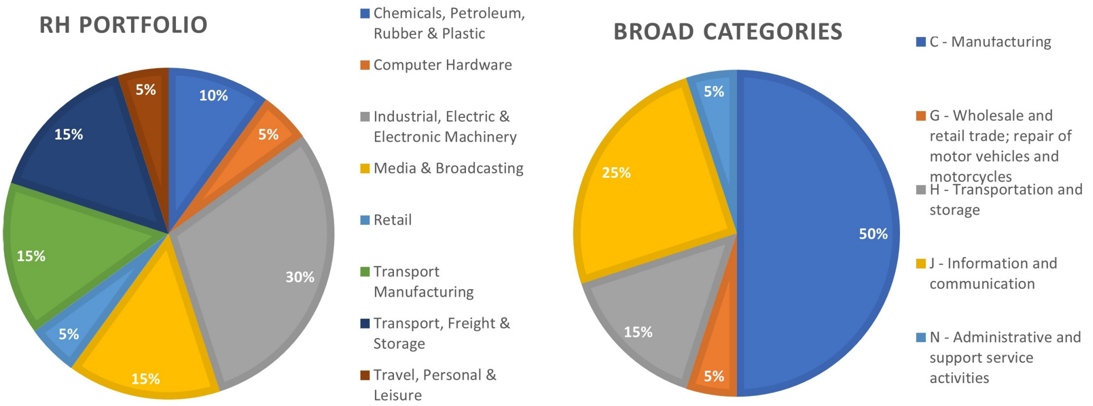

Figure 3.3: RH portfolio categories

3.4.1 Analysis and efficient frontier

The following section analyses the RH portfolio: first the performance of the RH portfolio is found, and then the portfolio composition and performance are compared with the optimal portfolios computed on the efficient frontier (the maximum Sharpe ratio and the minimum volatility).

The users' portfolio is represented as a portfolio that contains the 20 stocks that were in the Top10 popularity over the time horizon. Given that users have a small account size, it is reasonable to assume that they hold a limited number of stocks.

Following the procedure of Beck and Jaunin (2021), a proxy for portfolio weights can be estimated from the number of users holding a stock. The weights $w^R H$ of each stock i are calculated daily, as the fraction of users H holding the stock over the total number of users in that day t, the results are stored in a matrix of daily weights.

$$w_t^{RH}(i) = \frac{H_t(i)}{\sum_{i=1}^{I} H_t(i)}$$

From a matrix of prices of each stock, the daily returns R_t are calculated as the percentage change for each stock price P in each day:

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}} = \frac{P_t}{P_{t-1}} - 1$$

Two matrixes are obtained: one containing the weights assigned to each stock at each day, the other containing the returns for each stock and at each day. Then all stocks' weights are multiplied with their return for each day, and the daily total portfolio returns correspond to the sum of each stock return.

The metrics are calculated as follows:

The annualized return AR_t is calculated by multiplying the daily returns R_t with the number of open market days each year, which are 252:

$$AR_t = R_t \times 252$$

The return performance indicator for the portfolio is the average annual return AAR over the whole horizon T made of 584 days:

$$AAR = \frac{\sum_{t=1}^{T} AR_t}{T}$$

The performance indicator for the portfolio volatility is the annualized volatility AV, which is a measure indicating how risky is the portfolio:

$$AV = \sigma_{AR}$$

The overall portfolio performance is represented by the Sharpe ratio SR, which indicates the level of excess returns with respect to the risk (the risk free rate is assumed to be 0):

$$SR = \frac{AAR}{AV}$$

Efficient frontier:

The efficient frontier is found with the Monte Carlo analysis, a method that determines the performance of several portfolios by making different simulations in which each time random weights are assigned.

By plotting each portfolio as a point on a graph with return and volatility as axis, a threshold becomes visible and corresponds to the efficient frontier. The portfolios on the frontier are optimal and offer the highest return for a given risk (or the lowest risk for a given return), instead the portfolios under the frontier are sub-optimal because they provide a too low return with respect to the volatility. The reference portfolios are the ones that on the frontier have maximum Sharpe ratio and minimum volatility.

To find the optimal portfolios first the daily prices of all the 20 stocks are extracted from Yahoo Finance data, as for the RH portfolio, the returns R_t are calculated with the percentage transformation and are annualized by multiplying them with the days in which the market is open (252).

Then the average is calculated as AAR_i for each stock i (they are not aggregated yet, since they must be still multiplied by the random weights). The results are stored in the matrix of returns AAR containing each stock return.

$$AAR_i = \frac{\sum_{t=1}^{T} AR_t}{T}$$

From the returns R_t the matrix of the covariances is calculated between the stocks, and the covariance is annualized by multiplying it with the open market days, the obtained matrix COV will be used next with the weights to find the volatility.

The weights of the shares are set by generating a matrix W of positive numbers between 0 and 1 of dimensions $numbersimulations \times numberofshares$ ($10^5 \times 20$), where the sum of each row (i.e. weights of the portfolio) is 1.

The performance indicator for the return is found by multiplying the matrix of weights W by the average annualized returns AAR, from which a matrix AR is obtained, containing the average annual return for each of the 10^5 portfolios

$$AR = W \cdot AAR$$

The annualized volatility AV is the standard deviation calculated from the previously found weights and covariance matrix. w is the weight of a single portfolio from the matrix W and wt is the transposed of w.

$$AV = \sqrt{wt \cdot (COV \cdot w)}$$

The Sharpe ratio is as before

$$SR = \frac{AAR}{AV}$$

The optimal portfolios are selected among all the portfolios in the simulation as the one with the minimum volatility and the one with the maximum Sharpe ratio, their weights are then plot in the following bar chart Figure 3.5.

The plot in Figure 3.4 shows each portfolio as a point, and is characterized by the volatility and the return on the axes, and the Sharpe ratio indicated with the color (toward green for high Sharpe ratios), in the first graph also the single shares appear, in the second the optimal portfolios only.

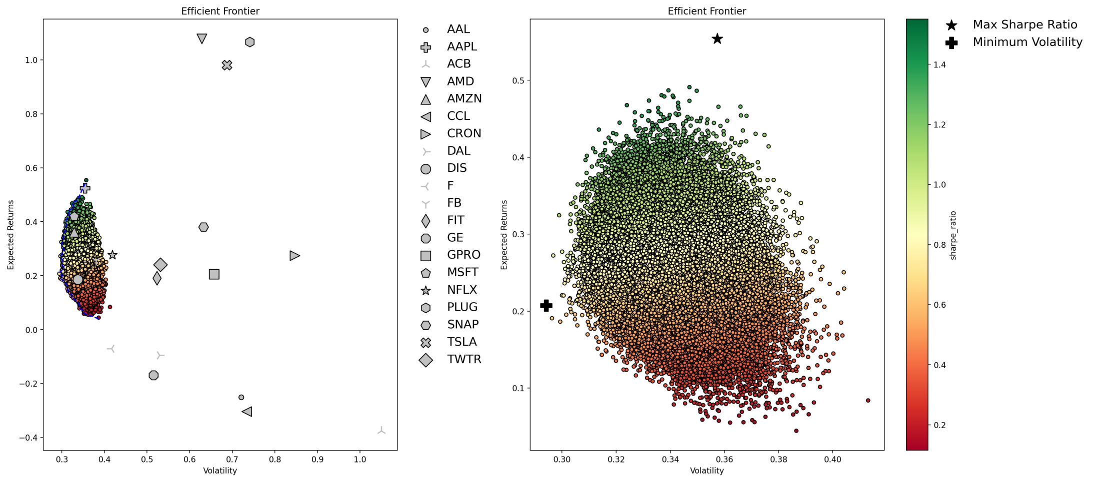

Figure 3.4: RH efficient frontier

PERFORMANCE	RH USERS	MAX SHARPE	MIN VOLATILITY
ANNUAL RETURNS	20.11%	55.39%	20.72%
ANNUAL VOLATILITY	41.34%	35.73%	29.42%
ANNUAL SHARPE RATIO	0.50	1.55	0.70

Table 3.3: Comparison between the RH and the optimal portfolios

The Sharpe ratio tells the portfolio performance with respect to the level of the volatility, and it is useful to compare the different portfolios and see if the return is adequate given the level of volatility and how much investors are 'exploiting' their portfolio potential, in this case the portfolio can potentially reach a 1.55 Sharpe ratio or a 30% volatility. The return of RH shares is similar to the optimal portfolio with minimum volatility, around 20%, however, the Sharpe ratio is lower (0.50 against 0.70), as the RH portfolio has higher volatility. If the higher volatility leads to higher returns and is therefore adequate, then the Sharpe ratio should be higher, but in this case, the higher volatility does not improve performance, and therefore, there is an excessive volatility that does not contribute to increase the returns. Indeed, from analyzing the weights in the following bar plot Figure 3.5, there are some shares that have a high weight in the RH portfolio and minimum weight both in the optimal portfolios maximum Sharpe ratio and minimum volatility, therefore for some reason, users hold large quantities of highly volatile shares that do not contribute to returns. It may again be an irrational behaviour induced by a psychological bias such as overconfidence or gambling, or also, the level of attention might be high for those stocks.

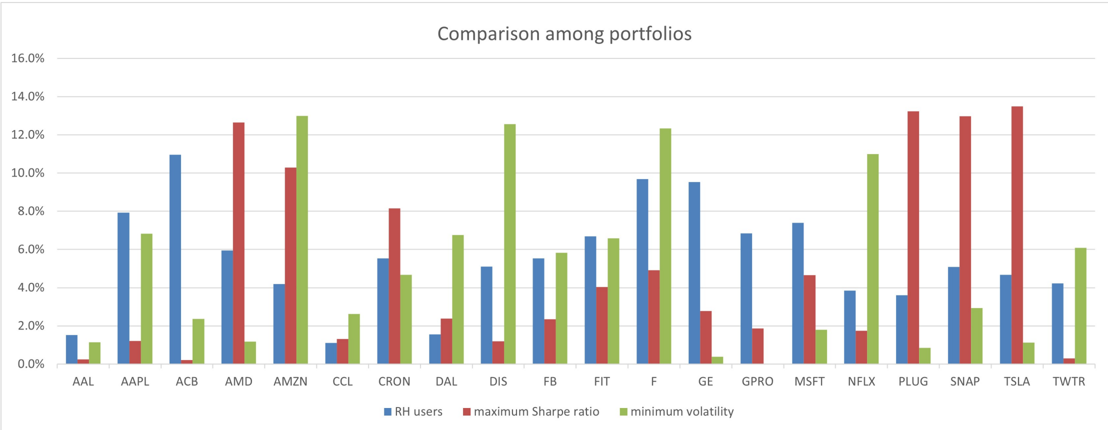

Figure 3.5: comparison of RH, maximum Sharpe ratio and minimum volatility portfolio's weights

Many of the stocks' weights are in between the maximum Sharpe ratio and the minimum volatility weights, in general they are closer to the latter. For the stocks ACB, GE, GPRO, there is an excessive investment, which increases the portfolio risk without improving the returns (and Sharpe ratio).

Chapter 4

Institutions

The analysed institutions were selected among the ones with largest Asset Under Management (AUM), for which data on the SEC database was available. The institutions belong to the following categories: ‘BG’ which indicates the three biggest institutional investors, having largest AUM according to Statista (Figure 4.1) and perform a wide range of activities; ‘BK’ stands for banks, ‘IN’ for insurance, ‘HF’ for hedge funds and ‘PF’ for pension funds.

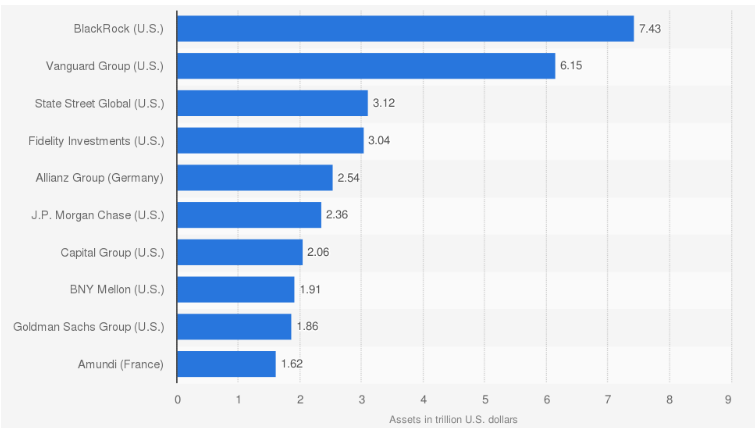

Figure 4.1: Largest AUM institutions. Source: Statista

Company name	Category	Symbol
BLACKROCK, INC	BG	BLK
VANGUARD GROUP INC	BG	Not listed
STATE STREET CORPORATION	BG	STT
CITIGROUP INC	BK	C
GOLDMAN SACHS GROUP, INC	BK	GS
MORGAN STANLEY	BK	MS
UBS GROUP AG	BK	UBSG
BRIDGEWATER ASSOCIATES LP	HF	Not listed
RENAISSANCE TECHNOLOGIES LLC	HF	Not listed
AQR CAPITAL MANAGEMENT LLC	HF	Not listed
MAN GROUP PLC	HF	EMG
TWO SIGMA INVESTMENTS	HF	Not listed
AVIVA PLC	IN	AV
BERKSHIRE HATHAWAY INC.	IN	BRK.B
PRUDENTIAL FINANCIAL INC.	IN	PRU
AMERICAN INTERNATIONAL GROUP INC.	IN	AIG
CALIFORNIA PUBLIC EMPLOYEES RETIREMENT SYSTEM	PF	Not listed
MICHIGAN LEGISLATIVE RETIREMENT SYSTEM	PF	Not listed
TEACHER RETIREMENT SYSTEM OF TEXAS	PF	Not listed

Table 4.1: Analysed institutions

The following analysis evaluates the institutions.
First the correlation is performed between each institution and between institutions and market data to evaluate common investing behaviours within each group (BG,BK,HF,IN,PF).
The relations with the market factors at the aggregate institutional level are evaluated with the regression for panel data, while the relation among the different groups are evaluated with the multiple linear regression. The last part compares the portfolio composition of the institutions and RH users.

4.1 Stock popularity

To find the most popular stocks among institutions the data about shares volumes were aggregated and sorted in descending order for each quarter and each institution, the first 10 stocks with highest volume were picked in each quarter and depicted with the bar plot. The figure shows the institutional top10 holdings at each quarter and can indicate how ‘diversified’ is the portfolio considering the ones with highest weights and how their volumes changes aver time.

If among the quarters the top10 stocks are more or less the same, the plot will be as the one on the left, with few stocks, indicating that the institution’s stocks with the highest weight in the portfolio tend to be the same over time, without many other additions. Instead, if the highest weight stocks tend to change more frequently over time, a figure as the left is seen, with a more populated bar plot.

The three biggest institutions together with insurance companies show a more static portfolio with respect to other institutions, since they hold the same stocks over 2 years’ time horizon, suggesting that they may follow a more ‘buy and hold’ and less diversified strategy, at least for their top10 stocks. In the last quarter of 2020, both the groups had increased significantly the shares of Apple Inc. and in general the BGs hold a very similar portfolio.

The popular stocks among the investors that are in common with RH investors are Apple Inc (AAPL), General Electric (GE), Microsoft (MSFT), Ford (F), Twitter (TWTR), Delta Airlines (DAL, only for Berkshire Hathaway); while the common popular stocks among the institutional investors are: Bank of America (BAC), Intel (INTL), Pfizer (PFE), Coca Cola (KO), AT and T (T).

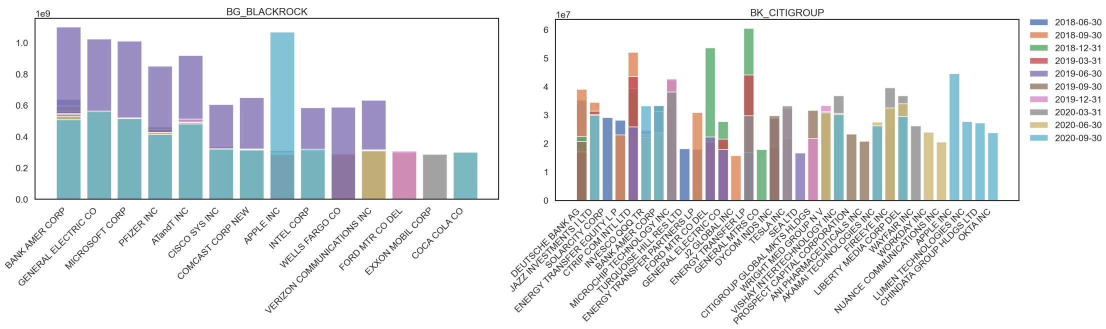

Figure 4.2: Top10 stock popularity on the institutions BG and BK

4.2 Correlations

A preliminary analysis among institutional shares' correlation with each stock shows that there are groups of institutions behaving in the same way.

For each stock the Spearman correlation is used to evaluate if institutions move together in buying or selling the stock. A qualitative assessment is made on the correlation matrix containing the correlation between the institutions, and it indicates if institutions belonging to the same category behave in the same way regarding the investment choices in the different quarters: positive correlations within the same group indicate concordance and homogeneity in the investing behaviour, while a negative correlation indicates discordance.

The BK group is the one with more available data, indicating it has most stocks in common with retail investors, also it is the most homogeneous one, in which institutions share the same strategy for many stocks, also the BG and HF share many stocks with retail investors, with HF and PF being most homogeneous after BK, and BG showing discordant behaviour for many stocks, instead IN is the group with less stocks in common with the retail investors.

The commonalities are also reflected in the behaviour respect to S&P500, VIX, market volumes and prices, but with more mixed results, indeed it is unlikely to see the same behaviour with respect to all the market variables.

Overall, BK and HF are more homogenous within the group regarding the investment activity, while BG and IN are more similar for the portfolio choices.

4.3 Regression

Institutional data is available with quarterly frequency, therefore there are only 10 time points in the two years horizon on which to perform the regression.

The data were transformed with the natural logarithm, and first, the regression for panel data was performed. In this case the single entities correspond to the institutions, and as before, the most suitable model is the fixed effect.

$shares_{jit}$ is the number of shares that the institution j holds for the stock i at time t, as before, the independent variables are $price_{it}$, VIX_t and $volume_{it}$, with $\gamma_j Dj_i$ capturing the individual factors.

$$shares_{jit} = \beta_0 + \beta_1 VIX_t + \beta_2 price_{it} + \beta_3 volume_{it} + \gamma_2 D2_i + \gamma_3 D3_i + u_{it}$$

It seems that price, VIX and volume could have an impact: as for RH the relation with price and volume is mainly positive, but contrary to RH the relation with VIX is mainly negative. R^2 is low, but higher with respect to the RH case.

Therefore, it looks that institutions are more attentive to market condition, are more risk adverse, and as the retail investors they invest when the market is moving.

Fixed effects regression for panel data					
	coefficient	T-statistic	P-value	R-squared	F-statistic
shares				0.142	0.000
const	-3.292	-3.281	0.001		
price_ln	0.645	12.834	0.000		
VIX_ln	-1.670	6.622	0.000		
volume_ln	0.999	20.165	0.000		

Table 4.2: Institutions regression for panel data

To investigate in more detail the investment behaviour of each institutional group, the OLS linear regression was performed on each institution, where $shares_{jit}$ is the share number of the institution j for stock i at time t and the independent variables are VIX, market volume and market price of stock i at time t:

$$shares_{jit} = \beta_0 + \beta_1 VIX_t + \beta_2 price_{it} + \beta_3 volume_{it} + u_{it}$$

The total number of observations is 360, and the following analysis is based on the statistics of the data. The level of significant cases, calculated as the average of significant values for price, VIX and volume over the total observations, is low, about 15%.

The data for each group are derived by isolating the significant cases and finding the percentage of positive and negative values among the significant ones for each group and each variable. Specifically, BK is the only group having a positive relation with VIX while HF is the only group with a significant prevalent negative relation with prices and volumes, BK and HF are therefore the groups showing some differences from the other institutions.

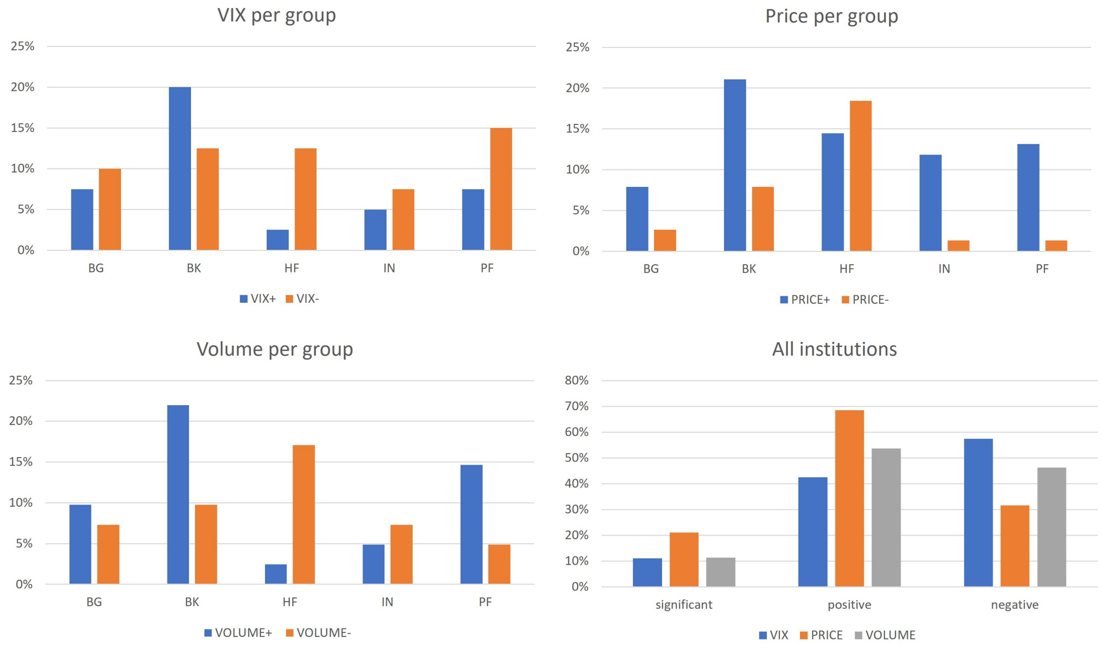

Figure 4.3: Multiple linear regression for institutions

4.4 Portfolio

From the aggregate institutional information on Whalewisdom, a website showing data about institutional investors and companies, it is noticed that the sectors in which institutional investors have more holdings are: finance, information technology with an increasing trend, consumer discretionary, consumer staples, energy with a decreasing trend and communication. A significant sector, finance, is not found in the RH portfolio.

The most viewed files on the website are in the table Table 4.3, given that the institutions use more sophisticated tools such as Bloomberg or Nasdaq, it is likely that it represents retail investors' interests in institutions.

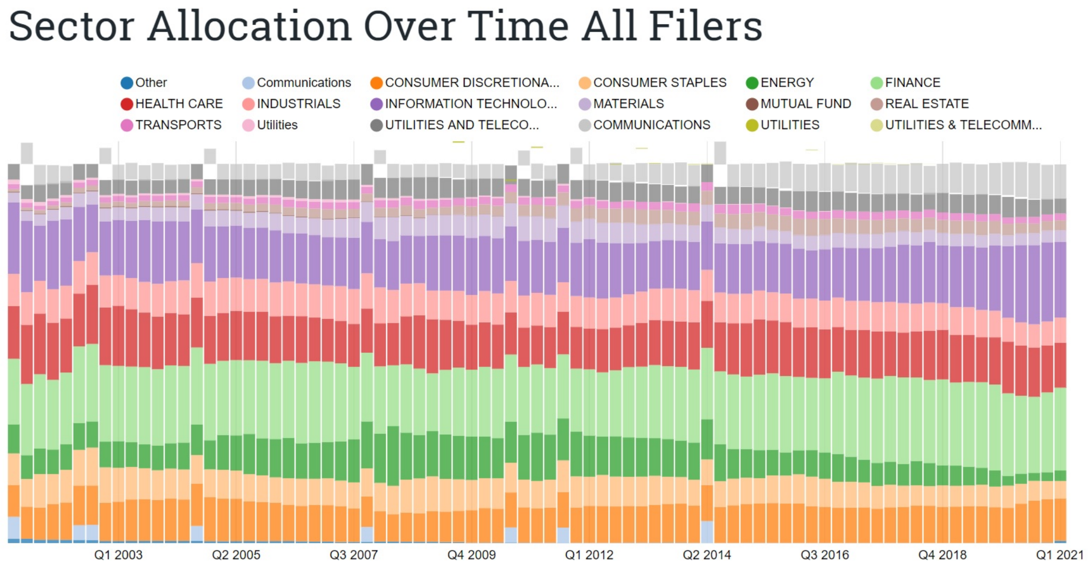

Figure 4.4: Institutional portfolio. Source: Whalewisdom

Filer Name	City
BERKSHIRE HATHAWAY INC	OMAHA
SCION ASSET MANAGEMENT, LLC	SARATOGA
MELVIN CAPITAL MANAGEMENT LP	NEW YORK
BLACKROCK INC.	NEW YORK
BRIDGEWATER ASSOCIATES, LP	WESTPORT
RENAISSANCE TECHNOLOGIES LLC	NEW YORK
ARK INVESTMENT MANAGEMENT LLC	NEW YORK
BAKER BROS. ADVISORS LP	NEW YORK
BAILLIE GIFFORD & COMPANY	EDINBURGH
TIGER GLOBAL MANAGEMENT LLC	NEW YORK

Table 4.3: Most viewed institutions on Whalewisdom website. Source: Whalewisdom

Chapter 5

Institutions and users

The following chapter compares RH users and institutions, first a regression for panel data is performed between users and institutions, then the analysis is performed on, instead of the RH popular stocks, on the popular institutional stocks, to evaluate if retail investors holding the institutional stocks behave differently and could be considered informed retail investors.

5.1 Regression - users and institutions

A panel data regression is performed to evaluate if there is a relation among users and institutional investors. The entities representing the dependent variable are the logarithm transformation of the user's shares $users\_ln$ for each stock i at time t and the independent variable is the logarithm of the shares' volume $shares$ of each institution . The regression type used is pooled, given that only volumes are compared and there are no individual effects to account for among the entities. The notation is as the fixed effect regression, excluding the individual terms.

$$users\_ln_{it} = \beta_0 + \beta_1 shares_{it} + u_{it}$$

The output shows that there is no relation between RH users and institutional investors since the independent variable is not significant and R-squared is almost zero.

Pooled regression for panel data

	coefficient	T-statistic	P-value	R-squared	F-statistic
users_ln				0.002	0.618
const	11.626	259.740	0.000		
shares	0.003	0.786	0.432		

Table 5.1: RH users and institution pooled regression for panel data

5.2 Stock popularity among institutions

The three biggest institution, BlackRock, Vanguard, State Street Corporation, have a similar portfolio, and BlackRock is among the most viewed institutional investors in the Whaleswisdom website, as shown in the Table 4.3, therefore they were taken as a reference for the most popular institutional portfolios.

The stocks belonging to the Top10 in each quarter of the two years horizon of the three institutions are the ones in Table 5.2, RH users behaviour on those stocks, excluding the ones already analysed (AAPL, F,GE,MSFT), looks similar to the RH top10, with a significant increase during the Covid outbreak as seen in Table A.4. As expected, there is a rapid increase of PFE in July 2020, when the company Pfizer Inc. showed promising results for the

Covid vaccines trials[1]. The stocks INTC, CMCSA, VZ and CSCO show an increasing trend without peaks. Further analysis, not covered in this work, could be made to evaluate if for stocks with a flat curve over the whole time horizon have historically low level of news or announcements and tend to not grab individuals' attention.

Company name	**Ticker symbol**	**Country ISO code**	**Sectors**
APPLE INC.	AAPL	US	Computer Hardware
AT&T INC.	T	US	Communications
BANK OF AMERICA CORPORATION	BAC	US	Business Services
CISCO SYSTEMS INC	CSCO	US	Communications
COCA-COLA COMPANY (THE)	KO	US	Food & Tobacco Manufacturing
COMCAST CORPORATION	CMCSA	US	Communications
EXXON MOBIL CORP	XOM	US	Chemicals, Petroleum, Rubber & Plastic
FORD MOTOR CO	F	US	Transport Manufacturing
GENERAL ELECTRIC COMPANY	GE	US	Transport Manufacturing
INTEL CORP	INTC	US	Industrial, Electric & Electronic Machinery
MICROSOFT CORPORATION	MSFT	US	Industrial, Electric & Electronic Machinery
PFIZER INC	PFE	US	Chemicals, Petroleum, Rubber & Plastic
VERIZON COMMUNICATIONS INC	VZ	US	Communications
WELLS FARGO & COMPANY	WFC	US	Business Services

Table 5.2: Top10 BG institutional portfolio

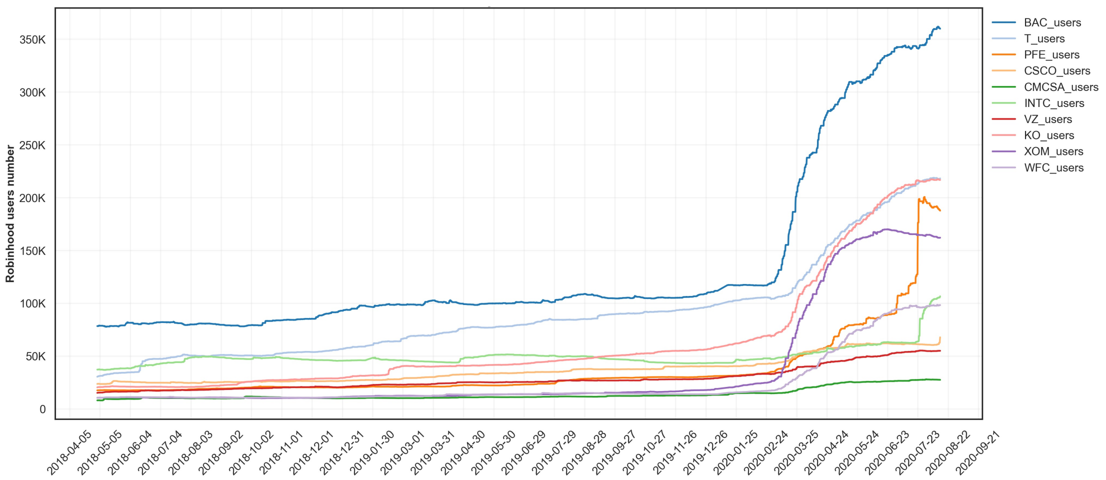

Figure 5.1: RH users and top10 institutional portfolio

[1]https://www.cnbc.com/2020/07/01/pfizer-stock-jumps-after-it-reports-positive-data-in-early-stage-coronavirus-vaccine-trial.html

5.3 Correlations

The Spearman correlation is performed as before among users and between users and market conditions. Among the users, the correlation is high indicating that there is a similar behaviour for all the stocks, which was not expected given that the majority of those is not part of the RH popular stocks.

The Spearman correlations among RH users and market conditions in Table A.4 show, as before, an increased positive relation with VIX and volumes in the FTH horizon respect to PCH, it is likely that RH users holding the institutional stocks are not more informed and behave under the same psychological biases as the RH users holding the most popular stocks on the RH platform.

5.4 Regression

With respect to the previous panel data with RH users holding the top10 RH stocks (Table 3.2), in the fixed effect regression for panel data of Table 5.3, where users hold the institutional stocks, the market conditions have more explanatory power, and the variables are all significant both in PCH and FTH. As before, the relation with VIX is negative in the PCH and positive in the FTH and the relation with volumes is positive in both horizons, conversely, here RH users have a negative relation with prices (mainly contrarian strategy) while with in the top10 stocks the relation was positive.

	full-time horizon				
	coefficient	T-statistic	P-value	R-squared	F-statistic
users_pct				0.182	0
const	-0.092	-12.002	0		
price_ln	-0.007	-6.956	0		
VIX_ln	0.004	10.483	0		
volume_ln	0.007	18.492	0		
	pre-Covid horizon				
	coefficient	T-statistic	P-value	R-squared	F-statistic
users_pct				0.069	0
const	-0.042	-9.299	0		
price_ln	-0.002	-2.122	0.034		
VIX_ln	-0.035	-8.408	0		

Table 5.3: RH and institutions regression for panel data

Regarding the multiple linear regression, on the aggregate the results are the same as the panel data and the previous top10 RH analysis.

In the FTH, the strategy is contrarian for VZ and momentum for BAC and CSC, and the average R-squared is lower respect to the PCH (0.1 against 0.21).

In the PCH, a contrarian strategy is followed for VZ,XOM,WFC and a momentum strategy for PFE, CMCSA, the stocks for which each strategy is followed have heterogeneous financial metrics (Figure A.2). The full regression values are found in the appendix Table A.5.

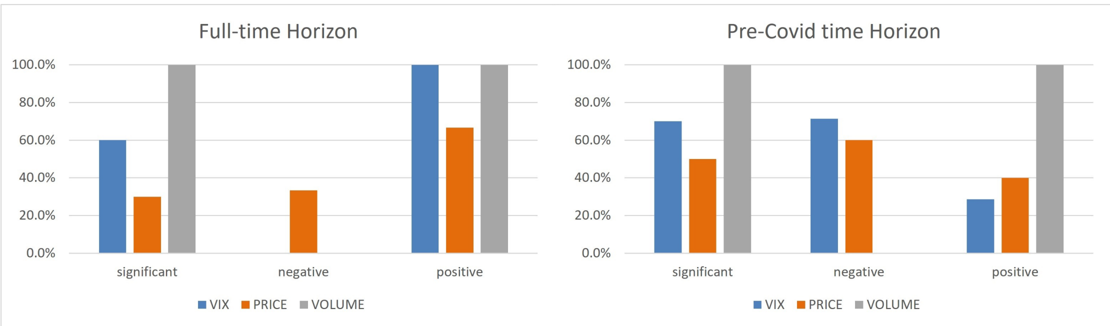

Figure 5.2: RH and institutions multiple linear regression

5.5 Portfolio

Regarding the portfolio categories of the most popular stocks, in the RH portfolio, the categories Communications, Business Services (i.e. financial) and Food&Tobacco are not present. Instead, RH portfolio contains Retail Transport, Freight&Storage, Media&Broadcasting, Travel&Personal Leisure which institutions hold but are not present in the Top10. The categories in common are: Chemicals, Petroleum, Rubber&Plastic, Computer Hardware, Transport Manufacturing, Industrial, Electric&Electronic Machinery.

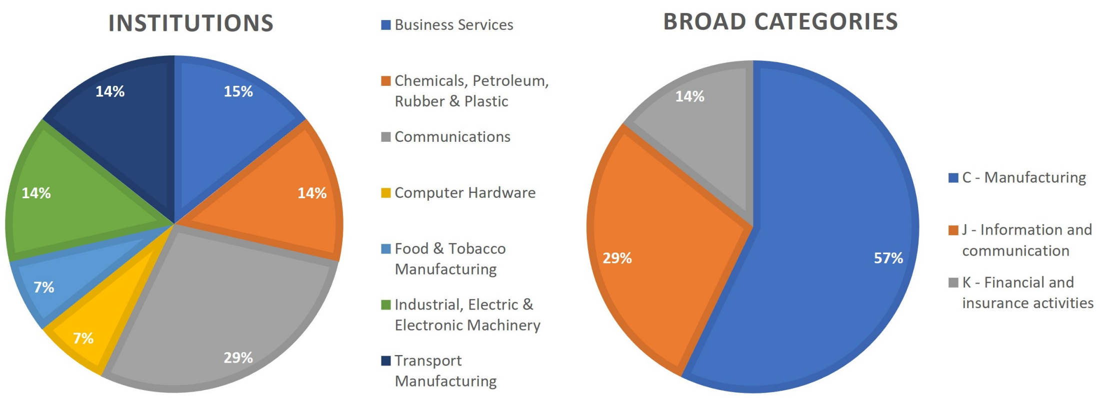

Figure 5.3: Top10 institutional portfolio categories

Efficient frontier:

As before, the efficient frontier is calculated considering all the most popular institutional stocks, the results of the Monte Carlo simulation in Figure 5.4 show the single stock performance (on the left) and the optimal portfolio location (on the right).

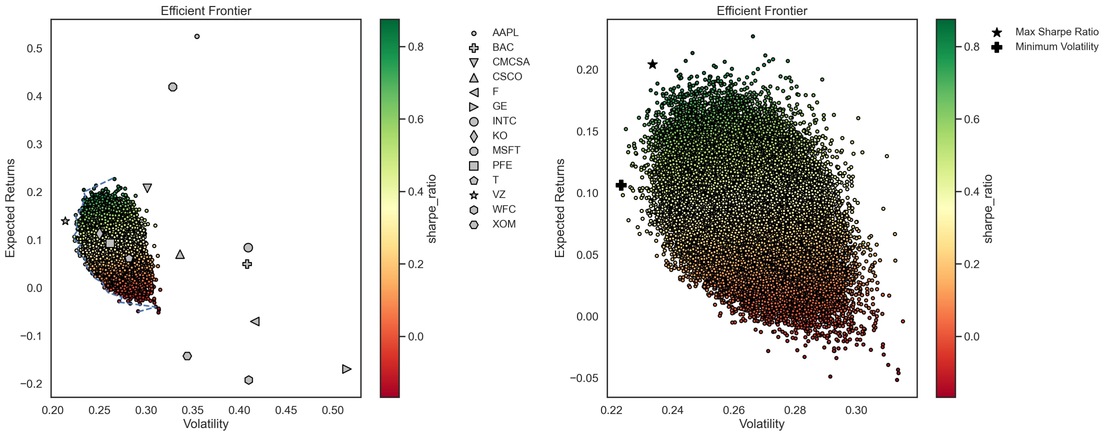

Figure 5.4: Top10 institutional portfolio efficient frontier

PERFORMANCE	RH USERS	MAX SHARPE	MIN VOLATILITY
ANNUAL RETURNS	11.23%	20.43%	10.63%
ANNUAL VOLATILITY	30.64%	23.34%	22.33%
ANNUAL SHARPE RATIO	0.37	0.88	0.48

Table 5.4: Comparison between the RH users holding the institutional portfolio and the optimal portfolios

Similarly to the top10 RH portfolio, returns are close to the minimum volatility portfolio, around 20%, but the risk is higher (30.64% against 22.33%) and provides a lower Sharpe ratio (0.37 against 0.48). Again, there is an excessive risk that does not contribute to increase the returns, indeed, as shown in the Figure 5.5 there is an excessive investment on stocks that have high volatility and do not increase the Sharpe ratio (F,GE,BAC). However, the returns are lower than the top10 RH portfolio and have and improved Sharpe ratio, due to the lower volatility. This portfolio has lower maximum Sharpe ratio, but also a lower minimum volatility, therefore, retail investors owning the institutional portfolio have a lower volatility portfolio that provides lower returns respect to the most popular RH stocks. However, this is due to the intrinsic characteristics of the stocks, the RH users behaviour shows as before some irrational patterns, by excessively investing in too high risk stocks relatively to the return they can provide.

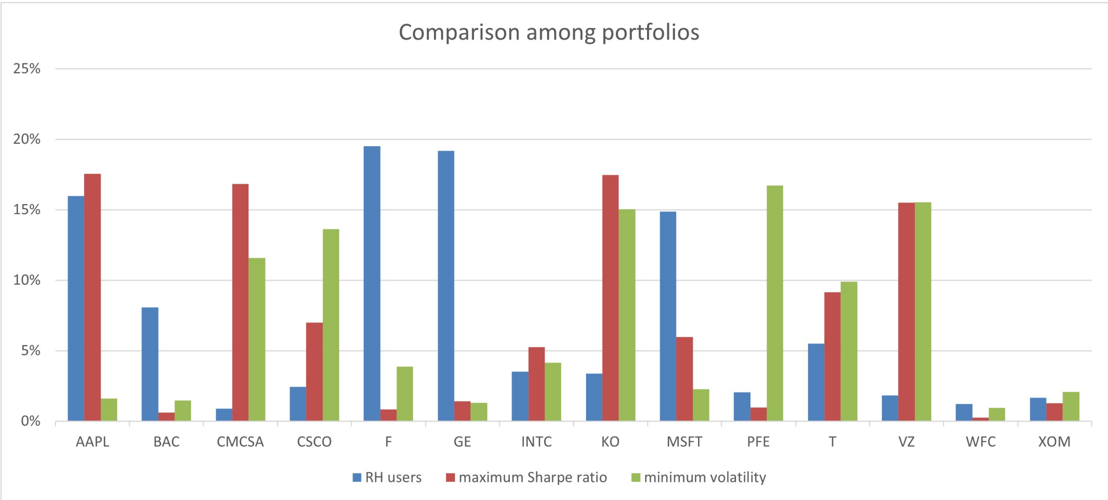

Figure 5.5: Comparison of RH users, maximum Sharpe ratio and minimum volatility with the institutional portfolio

Chapter 6

Conclusions

The study evaluates the relation between the investment activity and the market conditions S&P500, VIX, volumes and prices for both the groups institutions and RH users, and the relation is evaluated on both the RH's and the institutional most popular stocks.

RH investors are evaluated in two different time horizons, one that includes the Covid outbreak and one that excludes it, representing a market in normal conditions. It was found that RH users are positively related to VIX, volumes and prices when considering a highly volatile horizon, while in normal markets, the relations are weaker, negative with VIX and have less explanatory power. An important result is that retail investors invest in volatile markets and the relation with the market is even stronger when the volatility increases significantly.

Retail investors should not engage in financial activities that entail high risks, since they may not be able to bear the possible losses or they do not have proper tools and knowledge to effectively manage their investment, therefore, investing in such markets can be considered an irrational behaviour.

There is no evidence of a causal relationship between users and market factors, therefore it is not possible to conclude that the market volatility, volumes or prices are the causes of the investment decisions, also, if there was a causal relationship, it would be expected to have in both horizons a similar behaviour, with higher variables' significance and impact.

More likely indeed, is that it is not the market condition that has an influence on investors, but rather, a factor coincident with those specific conditions. Indeed, high market volumes and movements are associated with a higher level of news (Alanyali et al. 2013) and high level of news are associated to an increased level of attention toward financial markets and trading platforms (Aharon and Qadan 2020), therefore attention can be the real reason behind the irrational behavior of investing during volatile markets.

Interestingly, volatile markets did not only attract existing investors, but also new investors, who may have decided to place their first order in the financial markets right in a moment of very high volatility (indeed in the years 2018-2020 the RH platform experiences an increase of users from 6M to 20M, Table 2.2).

Regarding prices, RH users follow both a contrarian and momentum strategy, the first, which consists in selling well performing stocks and buying the bad performing ones, is followed for stocks with low financial performances. This kind of strategy might be driven by the psychological bias called 'disposition effect', for which investors sell the good stocks too soon, to secure profits and have an instant gratification, and instead keep to hold the bad ones to avoid realizing losses (Barber and Odean 2008, Shefrin and Statman 1985).

Another reason could be optimism and overconfidence for which investors are convinced of their beliefs that the market will go as they expect (Daniel and Hirshleifer 2015).

The RH behavior can be also explained by the demographics and the information sources: the users are indeed young, based in the US and are likely influenced by low quality sources such as social media or Youtube, which are among the visited websites of the users.

Institutional investors are instead risk averse, but they also show a positive relation with market volumes and prices, indeed they behave as rational investors and are not affected by the psychological biases since they have the necessary resources to make proper investigations and a defined strategy (Odean 1998). Banks and hedge funds are the most homogenous group and behave in a similar way within their group, also, they show for volumes, prices and VIX a different behavior with respect to the other institutions. Insurances and the largest AUM institutions have a similar portfolio and banks have most stocks in common with the RH investors. Since the institutional investor's activity is public, there might be a relation between the users and the institutions, however, the regression for panel data shows that there is no significant relationship. Regarding the portfolio analysis, RH investors have a low Sharpe ratio, that means the returns are low with respect to the risk they bear, and it is due to the fact that they hold an excessive amount of highly volatile stocks, that do not contribute to the returns, and therefore reduce the overall portfolio performance. This is another confirmation that retail investors are not risk averse and there are irrational factors driving their decision making. This emerging group may soon start to have an important role in the financial markets, on one side they can be beneficial and provide liquidity during market crisis (Beck and Jaunin 2021) while on the other side they can have a destabilizing role by increasing market volatility (Foucault et al. 2011; Baig et al. 2021; Eaton et al. 2021).

Appendix A

Appendix

2018-06-30		**2018-09-30**		**2018-12-31**	
symbol	users	symbol	users	symbol	users
AAPL	144336	AAPL	162376	AAPL	197624
GE	140955	FB	157820	F	186132
F	132586	F	154047	GE	184341
AMD	120454	GE	152075	FB	173900
MSFT	117209	AMD	142720	MSFT	161961
FIT	115804	MSFT	140153	AMD	159984
GPRO	113653	FIT	128417	CRON	157118
FB	103396	GPRO	121553	FIT	142406
TWTR	92933	TWTR	113254	GPRO	134048
SNAP	82999	NFLX	110628	ACB	131975

2019-03-31		**2019-06-30**		**2019-09-30**		**2019-12-31**	
symbol	users	symbol	users	symbol	users	symbol	users
ACB	283542	ACB	436807	ACB	511161	ACB	554160
GE	242430	GE	266625	GE	282046	F	304009
AAPL	234657	F	247895	F	272449	GE	296139
F	220335	AAPL	229174	MSFT	215454	FIT	253434
CRON	201550	MSFT	190044	AAPL	215104	GPRO	236044
MSFT	174195	CRON	187865	FIT	209850	MSFT	230109
FB	162049	FIT	175816	GPRO	189023	AAPL	205923
AMD	159951	GPRO	171705	CRON	181180	DIS	187647
GPRO	159024	AMD	161975	AMD	173160	CRON	181711
FIT	155912	TSLA	150314	TSLA	161671	AMD	165505

2020-03-31		**2020-06-30**		**2020-09-30**	
symbol	users	symbol	users	symbol	users
ACB	673801	F	818606	F	934520
F	403004	GE	713722	GE	843908
GE	349056	ACB	668612	AAL	653910
MSFT	310271	DIS	535842	DIS	618678
GPRO	296097	AAL	493400	DAL	583225
FIT	264400	DAL	463822	AAPL	567373
AAPL	253694	GPRO	456040	MSFT	554016
DIS	248138	MSFT	437304	CCL	488424
PLUG	206248	CCL	420505	GPRO	483842
SNAP	198199	AAPL	379191	TSLA	479885

Table A.1: Quarterly stock popularity on RH platform

Spearman rank correlation of RH users (full time horizon)

Symbol	SP500	VIX	volume	adjclose	returns	volatility
AAL	0.29	0.72	0.9	-0.95	0.01	0.76
AAPL	0.38	0.62	0.29	0.72	0.07	0.32
ACB	0.35	0.22	0.3	-0.48	-0.07	0.05
AMD	0.45	0.63	-0.01	0.9	-0.03	0.22
AMZN	0.44	0.42	0.1	0.79	0.03	0.24
CCL	0.32	0.72	0.9	-0.97	-0.01	0.77
CRON	0.31	0.38	0.25	0.29	-0.06	-0.08
DAL	0.3	0.69	0.84	-0.91	0	0.72
DIS	0.52	0.6	0.46	0.22	0	0.52
FB	0.09	0.53	0.4	0.04	0.03	0.41
FIT	0.59	0.51	0.2	0.17	0.01	-0.26
F	0.45	0.65	0.59	-0.77	0.04	0.6
GE	0.46	0.61	0.15	-0.73	0.01	0.45
GPRO	0.49	0.63	0	-0.65	0.02	0.27
MSFT	0.49	0.65	0.34	0.95	0.01	0.38
NFLX	0.3	0.29	0.09	0.45	-0.03	0.16
PLUG	0.52	0.56	0.68	0.81	0.06	0.48
SNAP	0.48	0.62	0.17	0.65	0.06	0.33
TSLA	0.53	0.5	0.32	0.66	0.08	0.24
TWTR	0.36	0.64	0.15	-0.35	0.03	0.27

Spearman rank correlation of RH users (pre-Covid time horizon)

Symbol	SP500	VIX	volume	adjclose	returns	volatility
AAL	0.58	0.12	0.11	-0.91	0.03	0.07
AAPL	0.28	0.25	0.13	0.13	-0.01	0.29
ACB	0.56	-0.3	0.28	-0.36	-0.09	-0.24
AMD	0.49	0.24	0.02	0.81	-0.08	0.07
AMZN	0.16	0.38	0.1	-0.03	-0.04	0.34
CCL	0.82	-0.1	0.21	-0.84	-0.03	0.24
CRON	0.32	0.27	0.32	0.6	-0.07	-0.12
DAL	0.65	-0.03	-0.09	0.47	0.01	-0.2
DIS	0.81	-0.14	0.13	0.93	-0.02	0.11
FB	-0.24	0.35	0.36	-0.7	-0.04	0.46
FIT	0.77	-0.07	0.2	-0.16	0.03	-0.03
F	0.7	0.01	0.07	-0.24	0.01	0.15
GE	0.6	0.03	-0.22	-0.36	0.07	0.05
GPRO	0.78	-0.09	0.21	-0.69	-0.02	-0.07
MSFT	0.73	0.04	-0.06	0.93	0.01	-0.08
NFLX	-0.23	0.33	0.39	-0.35	-0.08	0.54
PLUG	0.66	0.03	0.44	0.64	0.06	0.2
SNAP	0.68	0.01	0.09	0.51	0.05	0.1
TSLA	0.6	-0.07	0.13	-0.28	0.07	-0.08
TWTR	0.56	-0.2	0.08	-0.49	-0.03	0.13

Table A.2: Correlations between RH users and market conditions

Full-time horizon					Pre-Covid horizon			
company	VIX	PRICE	VOLUME	R^2	VIX	PRICE	VOLUME	R^2
AAL	0.0601***	0.0204***	-0.0004	0.3709	0.0012	-0.0015	0.0070***	0.069
AAPL	0.0015	0.0086***	0.0051***	0.2932	-0.0008	0.0025	0.0045***	0.0929
ACB	-0.0006	0.007	-0.0091	0.024	-0.0443	0.0106***	-0.0198*	0.036
AMD	-0.0027***	0.0040***	0.0035***	0.0858	-0.0031	0.0026***	0.0034***	0.0606
AMZN	0.0033***	0.0217***	0.0021*	0.3144	-0.0015	-0.002	0.0028**	0.0328
CCL	0.0472***	0.0375***	0.0144***	0.256	-0.0075*	-0.0114*	0.0233***	0.1383
CRON	-0.0061***	-0.0068***	0.0062***	0.2115	-0.0064***	-0.0075***	0.0065***	0.2139
DAL	0.0474***	0.0364***	0.0057***	0.44	0.0012	-0.0129**	0.0049***	0.116
DIS	0.0038***	-0.0022	0.0064***	0.3153	0.0004	0.0029	0.0063***	0.2251
F	0.0047***	-0.0009	0.0028***	0.2889	-0.0027***	-0.0080***	0.0028***	0.1674
FB	-0.001	0.0085***	0.0106**	0.1094	-0.0033	0.0188***	0.0125**	0.1156
FIT	0.0001	-0.0035***	0.0029***	0.1107	-0.0008	-0.0042***	0.0035***	0.1581
GE	0.0096***	0.0066***	0.0023***	0.4141	0.0012	0.0016	0.0028***	0.1915
GPRO	0.0022***	-0.0051***	0.0029***	0.3356	0.0022***	-0.0032***	0.0023***	0.2043
MSFT	-0.0028**	0.0071***	0.0052***	0.1658	-0.0057***	0.0045***	0.0045***	0.1227
NFLX	0.0021*	0.0214***	0.0076***	0.2251	-0.0038	0.0047	0.0062***	0.1259
PLUG	-0.0022**	-0.0037***	0.0029***	0.0369	-0.0053***	-0.0045*	0.0025***	0.0327
SNAP	0.0034***	-0.0021***	0.0047***	0.2373	-0.0012	-0.0023***	0.0045***	0.1921
TSLA	0.0022	0.0025	0.0077***	0.0972	-0.001	-0.0049	0.0076***	0.0482
TWTR	0.0031***	-0.0047**	0.0044***	0.2001	-0.0049***	0.0007	0.0045***	0.1648
AVG R2				0.22661				0.125395

Table A.3: RH multiple linear regression for each stock

Figure A.1: Top10 RH stocks financial metrics. Data source: Orbis

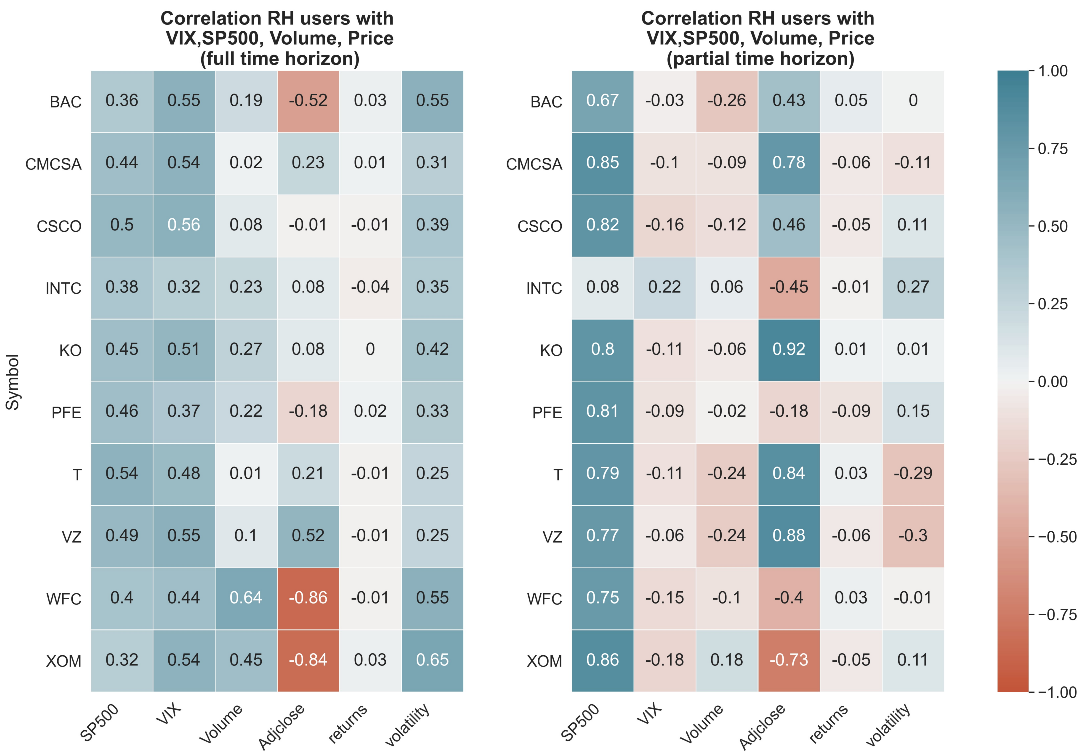

Table A.4: Correlations between RH users and market conditions on top10 institutional stocks

	Full-time Horizon				Pre-Covid Time Horizon			
company	VIX	PRICE	VOLUME	R^2	VIX	PRICE	VOLUME	R^2
BAC	0.0099***	0.0093**	0.0020**	0.3816	0.0037***	0.0053	0.0023***	0.0606
T	0.0018	-0.0007	0.0049***	0.0608	-0.0082***	0.0010	0.0049***	0.067
PFE	0.0009	0.0022	0.0046***	0.1484	-0.0024	0.0061***	0.0045***	0.0946
CSCO	0.0000	0.0112***	0.0126***	0.1928	-0.0046***	0.0031	0.0047***	0.0942
CMCSA	0.0052***	-0.0026	0.0047**	0.2307	-0.0012	0.0043**	0.0041*	0.0522
INTC	0.0076***	-0.0062	0.0099***	0.1337	-0.0021**	-0.0027	0.0032***	0.1313
VZ	0.0003	-0.0055***	0.0063**	0.1442	-0.0077***	-0.0051***	0.0073***	0.1656
KO	0.0013*	-0.0042	0.0026***	0.0602	-0.0034***	-0.0016	0.0027***	0.0405
XOM	0.0117***	0.0009	0.0040***	0.3775	0.0002	-0.0202***	0.0052***	0.0712
WFC	0.0103***	0.0009	0.0088***	0.4058	0.0063***	-0.0305***	0.0031***	0.1756
AVG R2				0.21357				0.09528

Table A.5: Multiple linear regression RH users holding the institutional portfolio

Figure A.2: Top10 institutional stocks financial metrics. Data source: Orbis

Bibliography

Aharon, David Y. and Mahmoud Qadan (2020). «When do retail investors pay attention to their trading platforms?» In: *The North American Journal of Economics and Finance* 53, p. 101209. ISSN: 10629408. DOI: 10.1016/j.najef.2020.101209. URL: https://linkinghub.elsevier.com/retrieve/pii/S1062940820301066 (cit. on pp. 2, 20, 49).

Alanyali, Merve, Helen Susannah Moat, and Tobias Preis (2013). «Quantifying the Relationship Between Financial News and the Stock Market». In: *Scientific Reports* 3.1, p. 3578. ISSN: 2045-2322. DOI: 10.1038/srep03578. URL: http://www.nature.com/articles/srep03578 (cit. on pp. 20, 24, 49).

Arkes, Hal R and Catherine Blumer (1985). «The psychology of sunk cost». In: *Organizational Behavior and Human Decision Processes* 35.1, pp. 124–140. ISSN: 07495978. DOI: 10.1016/0749-5978(85)90049-4. URL: https://linkinghub.elsevier.com/retrieve/pii/0749597885900494 (cit. on p. 5).

Audrino, Francesco, Fabio Sigrist, and Daniele Ballinari (2020). «The impact of sentiment and attention measures on stock market volatility». In: *International Journal of Forecasting* 36.2, pp. 334–357. ISSN: 01692070.

DOI: 10.1016/j.ijforecast.2019.05.010. URL: https://linkinghub.elsevier.com/retrieve/pii/S0169207019301645 (cit. on p. 2).

Baig, Ahmed, Benjamin M. Blau, Hassan A. Butt, and Awaid Yasin (2021). «Do Retail Traders Destabilize Financial Markets? An Investigation Surrounding the COVID-19 Pandemic». In: *SSRN Electronic Journal*. ISSN: 1556-5068. DOI: 10.2139/ssrn.3782159. URL: https://www.ssrn.com/abstract=3782159 (cit. on pp. 3, 51).

Ballinari, Daniele, Francesco Audrino, and Fabio Sigrist (2019). «When Does Attention Matter? The Effect of Investor Attention on Stock Market Volatility Around News Releases». In: *SSRN Electronic Journal*. ISSN: 1556-5068. DOI: 10.2139/ssrn.3506720. URL: https://www.ssrn.com/abstract=3506720 (cit. on p. 2).

Barber, Brad M. and Terrance Odean (2008). «All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors». In: *Review of Financial Studies* 21.2, pp. 785–818. ISSN: 0893-9454, 1465-7368. DOI: 10.1093/rfs/hhm079. URL: https://academic.oup.com/rfs/article-lookup/doi/10.1093/rfs/hhm079 (cit. on pp. iii, 1, 4, 50).

Beck, Philippe van der and Coralie Jaunin (2021). «The Equity Market Implications of the Retail Investment Boom». In: *SSRN Electronic Journal*. ISSN: 1556-5068. DOI: 10.2139/ssrn.3776421. URL: https://www.ssrn.com/abstract=3776421 (cit. on pp. 3, 10, 26, 51).

Daniel, Kent and David Hirshleifer (2015). «Overconfident Investors, Predictable Returns, and Excessive Trading». In: *Journal of Economic Perspectives* 29.4, pp. 61–88. ISSN: 0895-3309. DOI: 10.1257/jep.29.4.61.

URL: https://pubs.aeaweb.org/doi/10.1257/jep.29.4.61 (cit. on pp. 5, 50).

De Bondt, Werner FM and Richard Thaler (1985). «Does the stock market overreact?» In: *The Journal of finance* 40.3, pp. 793–805 (cit. on p. 25).

De Long, Bradford, Andrei Shleifer, Lawrence H. Summers, and Robert J. Waldmann (1990). «Noise Trader Risk in Financial Markets». In: 98.4, pp. 703–738. URL: https://www.jstor.org/stable/2937765 (cit. on p. 1).

Eaton, Gregory W., T. Clifton Green, Brian Roseman, and Yanbin Wu (2021). «Zero-Commission Individual Investors, High Frequency Traders, and Stock Market Quality». In: *SSRN Electronic Journal*. ISSN: 1556-5068. DOI: 10.2139/ssrn.3776874. URL: https://www.ssrn.com/abstract=3776874 (cit. on pp. 3, 51).

Foucault, Thierry, David Sraer, and David J Thesmar (2011). «Individual Investors and Volatility». In: *The Journal of Finance*, p. 39 (cit. on pp. 1, 51).

Gabaix, Xavier, Parameswaran Gopikrishnan, Vasiliki Plerou, and Eugene Stanley (2007). «A unified econophysics explanation for the power-law exponents of stock market activity». In: *Physica A: Statistical Mechanics and its Applications* 382.1. Applications of Physics in Financial Analysis, pp. 81–88. ISSN: 0378-4371. DOI: https://doi.org/10.1016/j.physa.2007.02.030. URL: https://www.sciencedirect.com/science/article/pii/S0378437107001392 (cit. on p. 1).

Jegadeesh, Narasimhan and Sheridan Titman (1993). «Returns to buying winners and selling losers: Implications for stock market efficiency». In: *The Journal of finance* 48.1, pp. 65–91 (cit. on p. 25).

Kahneman, Daniel (2011). *Thinking, fast and slow*. Macmillan (cit. on p. 3).

Korniotis, George M and Alok Kumar (2008). «Superior Information or a Psychological Bias? A Unified Framework with Cognitive Abilities Resolves Three Puzzles». In: p. 59 (cit. on p. 5).

Kumar, Alok (2009). «Who Gambles in the Stock Market?» In: *The Journal of Finance* 64.4, pp. 1889–1933. ISSN: 00221082. DOI: 10.1111/j.1540-6261.2009.01483.x. URL: http://doi.wiley.com/10.1111/j.1540-6261.2009.01483.x (cit. on p. 5).

Odean, Terrance (1998). «Are Investors Reluctant to Realize Their Losses?» In: *The Journal of Finance*, p. 24 (cit. on pp. 4, 18, 50).

Shefrin, Hersh and Meir Statman (1985). «The Disposition to Sell Winners Too Early and Ride Losers Too Long: Theory and Evidence». In: p. 15 (cit. on pp. 4, 18, 50).

Smales, L.A. (2021). «Investor attention and global market returns during the COVID-19 crisis». In: *International Review of Financial Analysis* 73, p. 101616. ISSN: 10575219. DOI: 10.1016/j.irfa.2020.101616. URL: https://linkinghub.elsevier.com/retrieve/pii/S1057521920302593 (cit. on p. 2).

Welch, Ivo (2020). *The Wisdom of the Robinhood Crowd*. w27866. Cambridge, MA: National Bureau of Economic Research, w27866. DOI: 10.3386/w27866. URL: http://www.nber.org/papers/w27866.pdf (cit. on p. 3).