UNIVERSITY OF LJUBLJANA
SCHOOL OF ECONOMICS AND BUSINESS

MASTER'S THESIS

THE IMPACT OF TWITTER AND REDDIT ON POWER OF RETAIL INVESTORS AND MOVEMENTS IN SPECIFIC STOCK PRICES

Ljubljana, January 2023 MARKO FURIĆ

AUTHORSHIP STATEMENT

The undersigned Marko Furić, a student at the University of Ljubljana, School of Economics and Business, (hereafter: SEB LU), author of this written final work of studies with the title The Impact of Twitter and Reddit on Power of Retail Investors and Movements in Specific Stock Prices, prepared under supervision of prof. Aljoša Valentinčič, PhD.

D E C L A R E

1. this written final work of studies to be based on the results of my own research;

2. the printed form of this written final work of studies to be identical to its electronic form;

3. the text of this written final work of studies to be language-edited and technically in adherence with the SEB LU's Technical Guidelines for Written Works, which means that I cited and / or quoted works and opinions of other authors in this written final work of studies in accordance with the SEB LU's Technical Guidelines for Written Works;

4. to be aware of the fact that plagiarism (in written or graphical form) is a criminal offence and can be prosecuted in accordance with the Criminal Code of the Republic of Slovenia;

5. to be aware of the consequences a proven plagiarism charge based on the this written final work could have for my status at the SEB LU in accordance with the relevant SEB LU Rules;

6. to have obtained all the necessary permits to use the data and works of other authors which are (in written or graphical form) referred to in this written final work of studies and to have clearly marked them;

7. to have acted in accordance with ethical principles during the preparation of this written final work of studies and to have, where necessary, obtained permission of the Ethics Committee;

8. my consent to use the electronic form of this written final work of studies for the detection of content similarity with other written works, using similarity detection software that is connected with the SEB LU Study Information System;

9. to transfer to the University of Ljubljana free of charge, non-exclusively, geographically and time-wise unlimited the right of saving this written final work of studies in the electronic form, the right of its reproduction, as well as the right of making this written final work of studies available to the public on the World Wide Web via the Repository of the University of Ljubljana;

10. my consent to publication of my personal data that are included in this written final work of studies and in this declaration, when this written final work of studies is published.

Ljubljana, _____ Author's signature: _____

(Month in words / Day / Year,
e. g. June 1st, 2012)

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

FF3 – Fama & French 3 factor model

P/B – Price to Book ratio

SMB – Small Minus Big

HML – High Minus Low

CAPM – Capital Asset Pricing Model

AR – Abnormal Return

RMB – renminbi

ROE – Return on Equity

EBT – Earnings Before Tax

EBIT – Earnings Before Interest and Taxes

EBITDA – Earnings Before Interest, Taxes, Depreciation, and Amortization

IPO – Initial Public Offering

ETF – Exchange Traded Fund

SEC – The Securities and Exchange Commission

INTRODUCTION

In January 2021 companies like GameStop and AMC experienced a sudden increase in their stock prices, and in the months prior they have been some of the most discussed stocks on social media platforms Twitter and Reddit (Long, Lucey, & Yarovaya, 2021). Exceptionally high amount of discussion on social media sites made a lot of people believe that retail investors have organized themselves on these platforms and are specifically targeting these companies to raise their stock price (Chohan, 2021). This led to my interest in the topic and desire to contribute to the discussion and raise the amount of available research.

In analyzing this topic, subject will be separated into three distinct parts that all influence one another and have to be analyzed to get a complete picture. These are retail investors, selected social media platforms, and selected stocks. As the first aspect, retail investors are defined as using Internet and social media/online platforms to inform themselves regarding the financial news and stock information, and they are the ones making the decision to invest and therefore participate in the discussion. In this thesis, research on investors will predominantly be focused on analyzing what are their investing habits and which platforms they use to gather information. Second part of the subject are the social media platforms that have been chosen for this research - Twitter and Reddit. Research shows that they contain the highest amount of discussion related to performance of companies on the stock market (Liew & Wang, 2016; Rothman, 2019). Reddit was the main focal point in late 2020 and early 2021 when GME and AMC experienced their sudden rise in stock price, and many news outlets referred to the people participating in this event as the 'Reddit Investors' (Long, Lucey, & Yarovaya, 2021). Twitter has been one of the most popular social networks since its inception, and resulting from that also has a high amount of financial discourse on its platform (Gu & Kurov, 2020). Together, these two platforms cover a large part of online discussion about the stock market and form an integral part of understanding how social media affects stock prices. In this thesis we will gather data from web scraping tools on how many times has a specific company been mentioned on Twitter and Reddit in a single day, for the past 3 years, and use this data for analysis of the companies. These companies that experienced a large change in their price are the third factor, and arguably the most important one. Thesis will analyze five popular companies by constructing Fama French 3 factor models to calculate by how much did selected stocks over/underperform their respective expectations. Afterwards, if they have a significant overperformance, it will be tested by calculating correlation and making regressions between aforementioned daily mentions on Twitter and Reddit, and daily returns that the stocks achieved in the specific period. By doing this we will be able to see how much of stock's returns can be explained and attributed to discussion on two selected networks, and what kind of correlation exists between the two.

The purpose of this research is to show the amount of correlation that five selected stocks have with corresponding daily amount of discussion about them on Twitter and Reddit. This will draw attention to how much power retail investors that are participating in discussion on two selected platforms have in making the stock price of a company rise or fall. If this impact is high, there could be repercussions and legal changes by the governments and its agencies to make the market less volatile. Potential discussion can arise on whether or not this type of behavior would be considered market manipulation and warrant legal consequences, since traders, independent of one another, have decided to invest into the same asset, with only underlying connection between them being participation in the same discussion on social media.

Aim is to contribute to the already present collection of research about how social media and capital markets are connected, by giving empirical evidence on connectedness between stocks and Twitter and Reddit. The outcome of the models constructed in this thesis can be used to make further assumptions on social media's impacts on other stocks and capital markets in general. Since social networks have become integral part of most people's lives and investors have become accustomed to use them to gather research on future investments, I believe it is important to analyze if this type of behavior gives retail investors additional leverage in beating the market, and how will this impact the stock market in the long term.

Judging from the fact that research already shows that in the past stock market has been affected by media articles for different publications (Tetlock, 2007), I assumed that social media will also impact it to some degree. Resulting from that the main research question of this thesis is set as:

"The stock market is affected by the social media, and the participants on social media hold a certain amount of power that can partially impact the movements of the market"

This research question is tested by calculating how the stock performed compared to its expectations through Fama French 3 factor models, and afterwards by determining the relation between previously determined excess return and daily mentions on social media through regressions and calculation of correlation.

1 INTERCONNECTEDNESS OF SOCIAL MEDIA AND CAPITAL MARKETS

1.1 Overview of previously done research

Many prior studies have focused on interconnectedness of capital markets (specifically the stock market), and social media. In their research, Azar & Lo (2016) analyzed if it is possible to predict how the stock market would react to Fed's Federal Open Market

Committee Meetings via analyzing Twitter feeds. They found that tweets do not contain direct information about asset prices, but if investors have enough leverage, information that was contained in the tweets could be used to build portfolios that are able to outperform a benchmark portfolio on several different dimensions. Also related to Twitter, Gu & Kurov (2020) argue that Twitter sentiment provides new information regarding three different aspects – quarterly earnings of a company, analysts' price targets, and analysts' recommendations. These three aspects put together account for one third of Twitter sentiment's predictive ability for stock returns. Liew and Wang (2016) argue that relationship between returns on first day of 300 IPOs that happened in period between 2013 and 2014, and corresponding discussion on Twitter show a positive correlation. A positive relation is exhibited from offering price to opening price, and a negative one from opening price to closing price. Their research concludes that Twitter discussion and sentiment matter for returns on the day of IPO, but the nature of this relationship is complex. Bollen, Mao, & Zeng (2011) indicate that there is 87.6% accuracy in predicting daily closing values of DJIA by using Twitter feeds to asses a general sentiment about the market.

In his research, Tetlock (2007) tried to quantitatively measure the correlation between the media and stock market. The author concluded that high pessimism in media forecasted a downward pressure on market prices, after which a return to previous price levels was observed. Tetlock's research however was primarily focused on news media for correlation, mainly Wall Street Journal, instead of social media sentiment. On the topic of online trading platforms, Dorfleitner & Scheckenbach (2021) showcase findings that there exists a negative relationship between overconfidence in trading on online platforms and returns that are achieved.

Additionally, Rothman (2019) studied the effects of social media (in particular Reddit, Twitter, and Telegram) on price and volume of Bitcoin. Author found that the total volume of tweets can predict variations in Bitcoin's price and volume, and that also Reddit and Telegram showed higher correlation than Twitter. In another research of similar sentiment, Tandon, Revankar, & Parihar (2021) predicted the values of Bitcoin with 96% accuracy using models based on tweets about cryptocurrencies.

As far as the newer research goes concerning GameStop, AMC and its connecting short squeeze event, Zaremba, Umar, & Yousaf (2021) demonstrate that a robust positive correlation exists between the price of GameStop stock and the performance of high short interest indices; and in a related research Ali, Umar, Gubareva, & Yousaf (2021) argue that put-call ratio and the short sales volume both positively influence GameStop's returns. In same research authors advise that policymakers and regulators should focus on continuous monitoring of investor groups on social media sites, since they have possibility of creating market inefficiencies. Additionally, Long, Lucey, & Yarovaya (2021) argue that sentiments expressed in comments on Reddit do influence the

GameStop's intraday stock price, with comments that indicate a sentiment of fear or sadness having a greater influence.

1.2 Current condition of the financial markets

As seen in the Figure 1, the S&P 500 index over the previous two years moved from a long-lasting bullish market that lasted until the beginning of 2022, to a downward sloping bearish market that continues until Q3 2022 (date of writing this). Reason for this change is a combination of many different factors that started to take effect around the beginning of 2022. Conflict between Russia and Ukraine caused disturbances in major world export lines, and supply of raw materials, natural gasses and food, leading to an increase in prices all over the world (Liadze, Macchiarelli, Mortimer-Lee, & Juanino, 2022). Secondly, inflation has been steadily rising since 2020 and the outbreak of Covid-19 virus. Beginning of the pandemic led the world into a state of emergency and many banks and companies had to adjust their business models to fit into the new normal. By the year 2022, this effect had not fully disappeared, and with the Fed's quantitative easing methods and printing new money throughout the years following the outbreak, and ECB's rather stagnant approach of dealing with the inflation, it has progressively developed into highest levels this decade (Eichengreen, 2022). This behavior by central banks proved to keep the economy rising during the Covid-19 outbreak and in 2021, but as soon as the aforementioned disturbances in the supply lines happened, the approach was not adequate any more.

Figure 1: Performance of SPDR S&P 500 ETF (SPY) in past 2 years

Source: MarketWatch (2022).

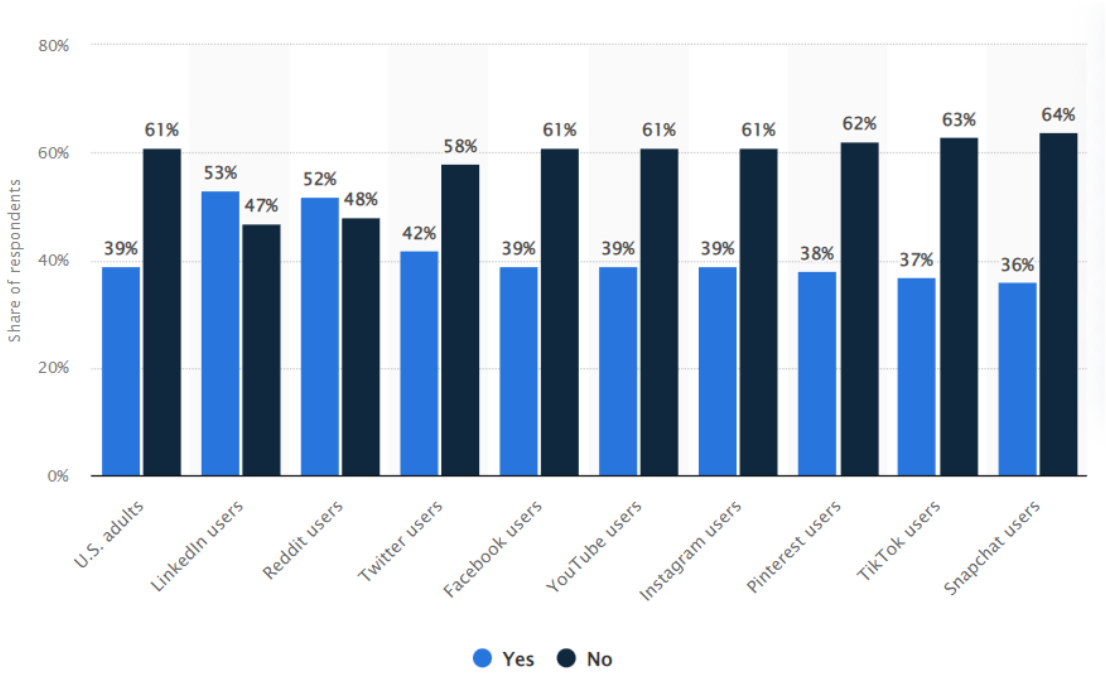
The year 2020 on the stock market was marked by a large fall in value of S&P 500 in March. The SPDR S&P 500 ETF fell from $337 to $228, and it has slowly been returning to pre-crisis values up until August. The remaining part of that year, as well as the whole 2021 saw a steady and constant increase in the value of the index, which lasted until December 2021 when the value started to decrease.

Over the course of those last 3 years two trends gained popularity, that are slowly shaping a new form of financial markets, with one of them being higher integration of stocks and its accompanying discussion into social media platforms, and second being the online brokerage platforms. Both of these aspects are a consequence of the lockdown of 2020

and early 2021, where people that already possessed financial/investment knowledge started to share it on different social media platforms, and communities that were focused on financial discussion started rapidly growing, with one of them being the Reddit's infamous WallStreetBets, which is discussed more in-depth in further chapters. With this new interest in the stock market, people that decided to invest, were primarily attracted to on-line, commission-free brokerage platforms which allowed them to start investing easily and without much additional expenses.

Regarding the actual number of users of social media sites that actively have money invested in the stock market, a YouGov (2021) survey, conducted on behalf of The Economist, questioned 1500 individuals from all different platforms on this topic. The results were following:

Figure 2: Share of social media users in the United States who have money invested in the stock market as of January 2021



Source: YouGov (2021).

As a baseline for comparison, the research showed that 39% of U.S. citizens have money invested in the stock market, while the other 61% does not. The only two social media sites where majority of the users own some investments in the stock market are LinkedIn and Reddit. LinkedIn having the highest proportion of users that participate in the stock market is due to it being a business-oriented network where investors will naturally gravitate towards it instead of the other, more entertainment focused platforms. Reddit, on the other hand, stands out with its 52% of sampled individuals owning investments on the stock market. Out of the ten most popular communities (subreddits) on Reddit none of them are directly connected to either finances or the stock market. The size of its

biggest financial oriented community – WallStreetBets is 'only' 12mn users, which compared to top ten where every subreddit has more than 25mn, and top three that all have over 35mn users, is relatively small (Baer, 2022), but still shows how big of an impact it has. Twitter users are on the third place with 42% of sampled users owning stocks, which is only slightly higher than percentages for the remaining platforms in the research, where the results all fall between 39 and 36%.

Regardless if the percentage of users that have investments in stocks is above 50% or around 40%, it is still a relatively high amount, and this high of a percentage can be attributed to internet's ability to provide fast access to information and an alternative to traditional brokerage accounts. In this case in a form of mobile apps that allow monitoring of bank accounts and investing into different financial instruments. These apps are generally categorized into 4 different types – apps for mobile payments, mobile banking apps, stock trading apps, and apps for managing cryptocurrencies and other crypto assets. Out of these four, the category most important for this research are the stock trading apps.

1.2.1 Online brokerage platforms

Online brokerage platforms operate on a zero-commission basis, and it is the core concept of their business. Retail investors can instantly place an order to buy or sell a share of any company that is being offered on the platform. In addition to that, one other aspect that helped these platforms gain popularity is availability to buy fractional shares, which is some cases can reach fractions of 0.0001 or lower. Since these brokerage platforms cannot earn money from commissions, they predominantly have a business model that is oriented towards Payment-for-order-flow (PFOF) to generate their income. The basis of this business model is that brokerage firm will receive a compensation for routing trades to a market maker for execution, so a percentage of profit made from trading is rerouted to the brokers (CFA Institute, 2016). Many of these platforms have a user interface similar to other social media apps where they allow investors to make portfolios and then share them with other users which can than like or dislike them. They also allow users to chat with each other and receive notifications on the stocks they invested in or marked as favorite. These are all markings and characteristics of sites like Facebook and Twitter, so the process of becoming retail investor is composed of just downloading an app and verifying identity, with no other barriers to entry. One additional advantage that makes young retail investors use them instead of traditional brokers are the low or zero commission fees, as opposed to high commissions and entry fees that are required to be paid to access traditional brokerage accounts.

These apps/brokerages have been massively gaining popularity over the last 6 to 7 years. As can be seen in the Table 1, the total amount of users of stock trading apps has more than tripled in period from 2016 to 2020.

Table 1: Stock trading app users from 2016 to 2020

Year	No. of users
2016	28.9 mn
2017	35.6 mn
2018	44.8 mn
2019	61.9 mn
2020	91.5 mn

Source: Curry (2022a).

For 2021 the information of usage statistics can be seen in the Table 2, from the research published by Curry (2022b) on behalf of the Sensor Tower, where the number of users for the top 12 most popular stock trading apps was compiled.

Table 2: Stock trading app users by app in 2021

App	No. of users (in mn)
Robinhood	22.5
eToro	13
Fidelity Investments	6.3
E-Trade	4.5
TD Ameritrade	3.8
CashApp	3
Charles Schwab	3
Plus500	2.3
WeBull	2
Trading 212	1.5
Degiro	1.2
Trade Republic	1

Source: Curry (2022b).

The most popular platform was Robinhood with 22.5 million users, and its closest peer is eToro with 13 million users. While they appear one next to another on this list, markets that they operate in are entirely different. Robinhood is a U.S. based brokerage, and not available in Europe and rest of the world. eToro, on the other hand, is entirely based in Europe with no coverage in the United States (Curry, 2022c). After the two leaders in the category of mobile brokerage platforms, the rest are either aiming for different types of investors, or only in the infancy of their stock trading business. Fidelity, TD Ameritrade and Charles Schwab are established brokers whose entry into mobile market has been different than the first two. Their business model was oriented towards senior investors, indicated by both higher commission fees, and by having a higher minimum investment amount that the broker is willing to take, which is not the case for Robinhood and eToro.

Other companies on the list differentiate themselves by being an all-in-one app, meaning that they include all of the four categories that were mentioned previously – mobile banking, payments, stock trading, and crypto trading. Main ones in this category are Cash App, Degiro, and also Revolut and Binance which were not mentioned on the above list. Plus500, WeBull, Trading 212 and Trade Republic are all new platforms that seek to take a portion of market share from Robinhood and eToro, by having a better user experience and by offering a larger amount of securities to invest into.

According to statistics published by Bevan (2022) and Curry (2022c) a steady rise in users can be observed over the course of the last 7 years for the two biggest online stock trading platforms.

Table 3: Robinhood and eToro user statistics (2015-2021)

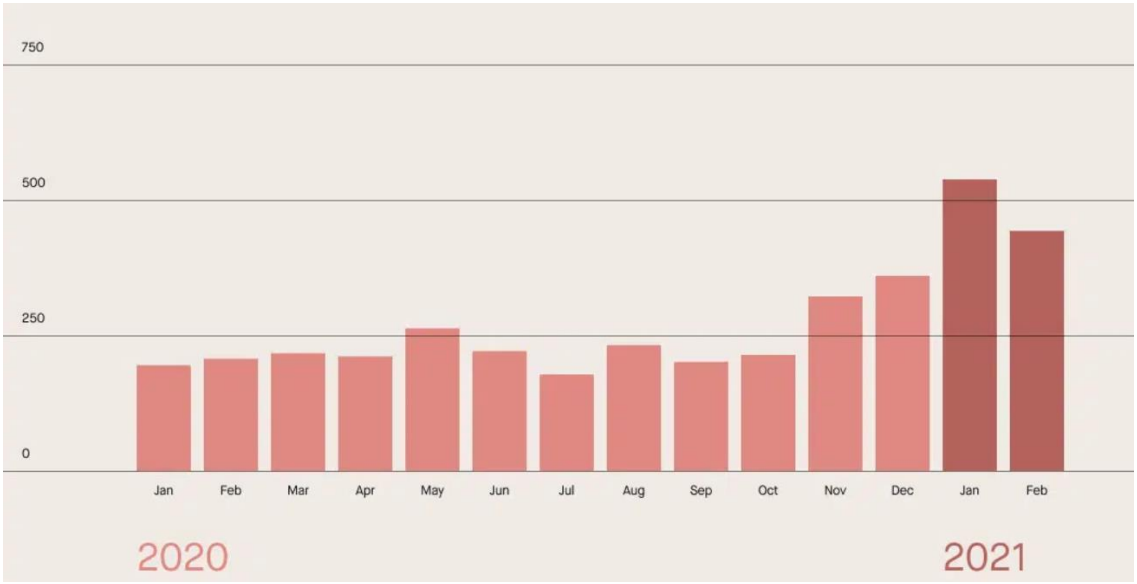
Year	Robinhood	eToro
2015	0.5 mn	4.7 mn
2016	1 mn	5.6 mn
2017	2 mn	8.1 mn
2018	6 mn	10 mn
2019	10 mn	12 mn
2020	13 mn	17 mn
2021	22.5 mn	27 mn

Source: Bevan (2022), Curry (2022c).

In 2015, 2 years after the launch of Robinhood and 8 years after the launch of eToro, platforms had 0.5mn and 4.7mn users respectively. This clear lead that eToro had comes from its early launch date of 2007, but as seen in Table 3, it had much harder time attracting newcomers in the second half of 2010s, up until 2021, which was a pivotal year for these platforms. During that year eToro recorded a 41% increase in their user base, while for Robinhood it was 73%. A quarter-by-quarter look at Robinhood user gains during 2021 and early 2022 shows that in Q1 2021 they had 17.3mn users, that increased to 21.3mn in Q2, which would make sense following the events of early 2021. But afterwards it underwent a volatile period where active user count fell to 18.9mn in Q3 and rose up again to 22.5mn. This would indicate that many people decided to try it after hearing about the GameStop rise in stock price and wanted to be a part of it, but as soon as the event started to lose media attention, they closed their accounts. In addition to that, at the time, many users started to protest against the platform by shutting down their accounts for blocking trades of GME, AMC and NOK among others (Lambert, 2021).

By looking at the statistics published by Robinhood and other platforms about their business model and profitability, we see what kind of investors are primarily using these platforms and their characteristics. The average transaction size on Robinhood over the course of 2020 has been between $200 and $500, with it staying below $250 for the large majority of 2020 and only rising to $500 as 2021 started.

Figure 3: Average transaction size on Robinhood through 2020 and beginning of 2021



Source: Bevan (2022).

From this we can estimate the average salary of people that invest through Robinhood. According to the experts' advice, 10% of after-tax income should be taken for investing into stocks (Jackson, 2022) and by making an estimate from the above figure that an average monthly investment on Robinhood equals $270, from this it can estimated that an average salary of a person that has investments in Robinhood is around $2700 after tax, or around $33,000 net per year. This falls below the median annual U.S. personal income which was $35,000 in 2019 (Kopestinsky, 2022), while higher than the average EU annual net salary which was around EUR 14,000 in 2020 (Eurostat, 2022). While this presumes that investor only makes one transaction per month, and thus is likely underestimating the value of a person's income, assumption is that an Robinhood investor is earning below average income. This assumption is further supported by a comparison of average account size between Robinhood, E-Trade, TD Ameritrade and Charles Schwab. Account size of an average Robinhood user is valued to approximately $3,500 while for the three remaining brokerages account sizes are all above or equal $100,000 (Nasli, 2022). $3,500 is roughly equal to above estimated $270 monthly payments over a year period, which would also indicate that accounts on Robinhood are much younger or the monthly payments are even lower than the above estimations. This further strengthens the assumption that the period from the beginning of Covid-19 until 2022 was vital for online brokerages, and the influx of new users that was observed had a high impact on their business model. Assets under management (or the total market value of all investments a company holds) of each of these companies correlate with the above figures, with Robinhood being the lowest with $80bn, then E-Trade with $600bn, TD Ameritrade with $1.3tr, and Charles Schwab at the top with $3.8tr (Nasli, 2022).

1.2.2 Summary of events of early 2021

During the course of 2020, investors on social media noticed that companies like GameStop, AMC, Nokia, and Blackberry have become some of the most widely shorted companies in the United States. These short positions were primarily held by institutional investors due to aforementioned companies having a declining stock price over the past years. These companies have once held a strong position in their respective industries, but over the years they started to lose their market share and their profits were declining, while Covid-19 pandemic even further blocked them from achieving their business goals. Stock being shorted is far from an irregularity, and the average short positions range from 1.6% to 2.9% of the outstanding shares, but GME and others stood out from this by having the percentage of shares sold short in relation to all publicly available shares be close to 100% or even above this figure after 2019 (Mackintosh, 2022).

Investors that were primarily gathered on Reddit's WallStreetBets community have noticed a theoretical possibility of a short squeeze happening to these stocks. Short squeeze is an uncommon event on the stock market where investors that have shorted the stock are forced to buy back shares they have borrowed at a higher price in order to limit their losses from increasing any further (Chohan, 2021). This would trigger an exponential price increase of that company's shares, and a potential for high profit for investors that have bought the stock while the price was low. The low likelihood of an event like this happening is due to the fundamentals of the companies involved; they are either near bankruptcy, unprofitable for a long time, or generally having bad business practices. Investors will be highly skeptical to invest in them, since their value is forecasted to only decrease, so the probability of a high number of investors suddenly changing their mind is very low.

Since many Reddit investors have grown up using services and goods from these companies, they have a sentimental and nostalgic feelings towards them, and felt connection and responsibility for them. In order to financially hurt institutional investors that are in their eyes being greedy and too aggressive, they had decided to be the reason behind the price increase that would trigger the short squeeze event. This reasoning was not purely altruistic since these smaller investors also had an incentive due to buying these stocks at a very low price and selling it once it has hit its peak, would make large returns on their investment. Long positions that retail investors were taking on these specific companies did not have any other reason or backing behind them besides the two already mentioned, since the prospects of these companies and their long-term value was still uncertain, and likelihood of them becoming profitable was low.

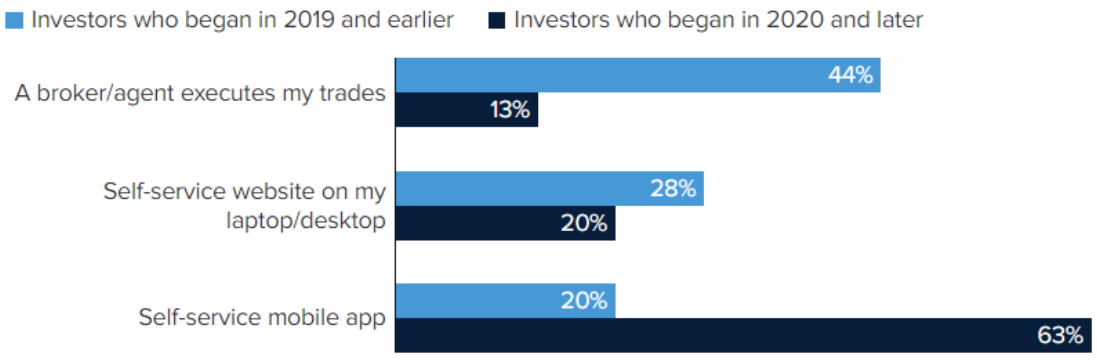
Once the rise in price did happen (see Figure 7) both GME and AMC reached all time high stock price. As the price was increasing during January and February of 2021, many of the aforementioned online stock trading platforms blocked trading of related stocks

(Mackintosh, 2022) making investors unable to sell, and leading many to question the legitimacy and legality of these actions and trustworthiness of the platforms.

1.3 Social media as the mediator of information

Previously mentioned integration of social media and stock market, and new investing platforms can be difficult for older investors to learn and adapt to. This can lead to a division between generations of investors, where older ones are more orientated towards older type of media, like newspapers and TV network channels, while younger investors will tend to use type of media that they are more familiar with. According to Clor-Proell, Guggenmos, & Rennekamp (2020), 65 percent of investors aged between 35 and 54 that own online investing accounts believe that being able to use apps on their phone to monitor and conduct investments is essential, and the numbers are even higher with younger investors aged between 25 and 34 where 72 percent use apps to track their investments. The same research finds that push notifications on apps for online investing result in more favorable investment judgments, but at the same time, investors will have a higher fear of missing out, and be more sensitive to the effects of the push notifications. Additionally, a survey published by Rodriguez (2021) on behalf of CNBC, analyzed differences between different age groups, ethnicities, genders, income and education groups, and how they gather information and invest their income. In a survey, 5,523 U.S. citizens were questioned, with sample of people being made to adequately represent the U.S. population through age, race, sex, education, and geography using the Census Bureau's American Community Survey.

Figure 4: Tools used for investing



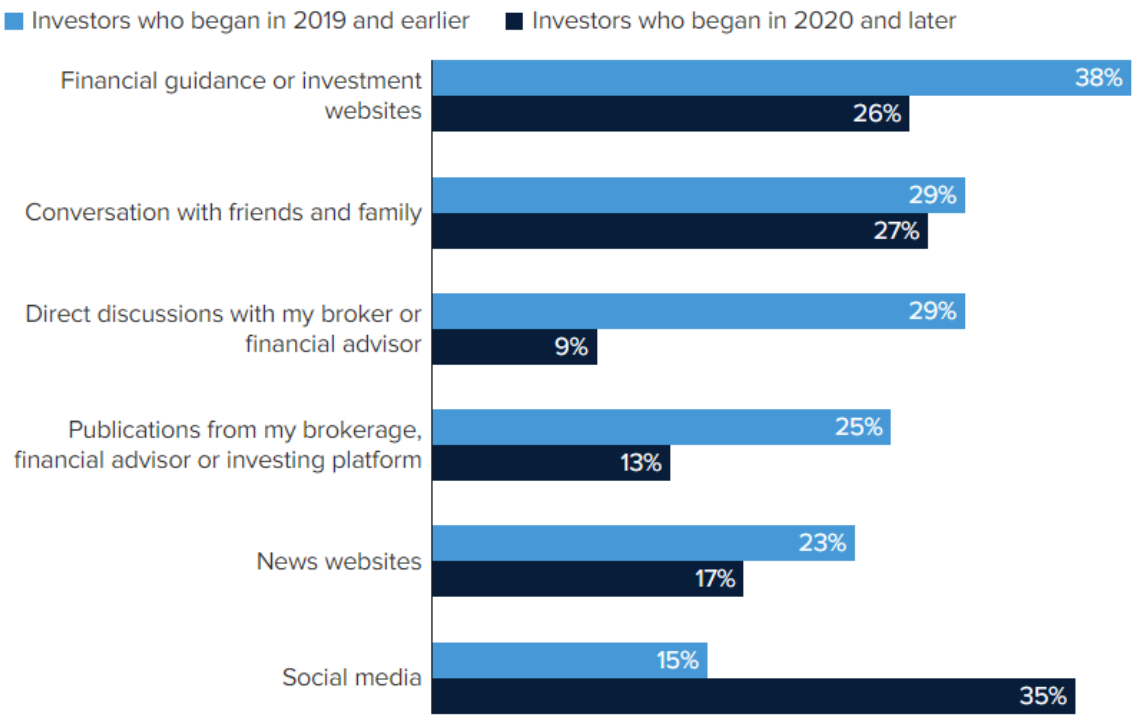
Source: Rodriguez (2021).

Investors who began in 2020 or later are using online brokerage apps for trading far more than typical agents or websites. While investors who started before 2020 are doing the exact opposite, with majority executing their trades through an agent or a self-service website. Many of the 2020 onward investors have also been faced with only a bull market that was happening from the Covid-19 impact in March of 2020 until the second half of 2022. This coupled together with the assumption that they are relatively new,

inexperienced and without proper education on the matter, can lead to a false sense of security and confidence that their trading strategies are better than those of the others. Mercer, Kadous, & Zhou (2022) argued that investors rely on the low-quality advice from social media because they believe it to be predictive of future returns. They argue that low quality advice that is received from social media is influential even when investors know that they should not rely on it, since rejecting a piece of information is a two-step process. First it is automatically accepted when initially received and then by systematic processing it is evaluated and rejected if deemed adequate. If individuals do not realize that these advices are affecting them, they may fail to engage in the second step, and thus "mentally contaminating" their judgments.

In their research on the internet as an information intermediary, Drake, Thornock, & Twedt (2017) classified sources of information for investors into three different categories – professional, semi-professional and non-professional. Social media sites along with blogs and forums have been sorted into the non-professional category, with a slight overlap with the semi-professional. According to research, this indicates that information obtained on these websites is of lower quality, and that it is not posted by professional intermediaries. Opposite of that, professional category provides information disseminated via the Dow Jones Newswires, and the semi-professional includes both paid individuals, business news websites and investment news websites. To complement Figure 4, additional statistics on how respondents collect their information and research new investment ideas are presented within the same survey. Results are broken down into six different categories, and two different groups of investors.

Figure 5: Share of survey respondents who say they use the following to research investment ideas



Source: Rodriguez (2021).

There is a clear distinction between investors who began before and after 2020, social media being the biggest difference between the two, with it being the source of information for 35% of 'new' investors, and for only 15% of 'old' investors. Second most popular source for younger investors are financial websites, which are tightly connected with social media since every publication typically has a social media account on which it posts news and shares links to the original website. These types of investors spend the least amount of time getting information directly from brokers or brokerage platforms since, as disclosed in Figure 5, majority of them do not use that type of service, but rather a self-service platform where they have no direct contact with the agent/broker. On the other hand, investors that started before 2020 are more conservative when it comes to getting information from social media with only 15% of them utilizing it. They prefer established sources like discussions with the brokers, and publications from brokerages that they are invested in, and by biggest margin financial guidance and investment websites.

Additional fact that helps social media platforms transmit information to wider audience is that according to the Rennekamp & Witz (2021), who on their research on linguistic formality and audience engagement concluded that individual investors are more sensitive and responsive to new information that uses informal language. Their findings also conclude that when combined with higher audience engagement, informal language can lead into better responsiveness to invest into the specific company. Social media sites

benefit greatly from this since casual tone and writing style is widely prevalent on all of them, and is according to this research one of the factors why social media is an efficient mediator of information on stocks. This also aids the companies themselves that use social networking platforms to better connect with possible investors. More specifically, Elliot, Grant, & Hodge (2018) provided evidence that when CEOs use Twitter to communicate information and news to investors, it forms a bond between two parties and makes a stronger connection. It makes investors trust the company more and negative news or earnings report have lower impact than they otherwise would. This shows that social media has the ability to change the way information is perceived, and as a consequence of that, has an impact on stock prices in an indirect way.

Information shared on social media is of questionable quality, which has been indicated by the SEC on numerous occasions (SEC, 2020a; SEC, 2020b; SEC, 2021). The SEC warned that social media has for a long time been a place where "fraudsters" are spreading misinformation, conducting market manipulations and profiting at the individual investor's expense. This creates a dangerous field for investors to pick between reliable sources and what information not to trust, especially if they are not skilled enough to differentiate between the two.

2 POPULAR STOCKS ON SOCIAL MEDIA

2.1 Most talked about stocks on social media in 2021

In order to determine what were the most popular stocks on social media during 2021, multiple popularity lists will be analyzed that were published by different publications, and condensed into one comprehensive table, from which we will be able to determine similarities and characteristics that these stocks possess. Publications included will be: Business Insider, CityIndex, Investor Place, Forex, MarketWatch, U.S. News, Yahoo Finance, Nasdaq, Motley Fool, and Fortune. These lists have different metrics and requirements by which they sort stocks, which will ensure that the outlook of popular stocks is not too narrow and in only one single aspect, but rather an all-encompassing view of general popularity of stocks on social media.

As seen in the Table 4, four publications had their lists in top 10 format, two in top 8, and four in top 5. All lists were published during the course of 2021 at different points, which explains the large number of different stocks.

Table 4: Lists of most popular stocks on social media in 2021 from 10 selected publications

Publication	Business Insider	CityIndex	Investor Place	Forex	Market Watch	U.S. News	Yahoo Finance	Nasdaq	Fortune	Motley Fool
1	TSLA	BB	CLOV	TSLA	NOK	GME	GME	CLNE	AMC	EXPR
2	LCID	AMC	BB	SDC	AMC	TSLA	AMC	IVR	GME	NAKD
3	ZM	GME	GME	PLTR	GME	AMC	UPST	TLRY	BB	AMC
4	SOFI	TLRY	AMC	MIR	BB	TLRY	CAR	CLOV	WISH	SNDL
5	AMC	NOK	CLNE	GME	TLRY	AAPL	HUT	BBBY	CLF	BB
6	GME	SNDL	WEN	BABA	BBBY	AMD				
7	AAPL	CLOV	WKHS	LCID	MVIS	NVDA				
8	WISH	BBBY	ASO	AMD	KOSS	BB				
9	TGT	PLTR	TLRY	GOEV						
10	FB	RKT	NOK	AMC						

Source: Fox (2022), Warner (2021a), White (2021), Warner (2021b), Van Doorn (2021), Duggan (2022), Kabir (2021), Amos (2021), Morris (2021), Tenebruso (2021).

Two most popular stocks during 2021 were AMC and GME, appearing in 9 and 8 lists respectively. These two stocks have been at the forefront of discussion in 2021, not only on social media, but in mainstream news outlets as well, receiving wide coverage and being one of the main financial news in the beginning of the year (Hasso, Müller, Pelster, & Warkulat, 2022). Afterwards, TLRY appeared in 5 lists. Tilray is a global pharmaceutical company, focused on medical cannabis and its therapeutic potential. On 15/12/2020, it conducted a reverse acquisition with Aphria (another company in the pharmaceutical industry, that was already publicly listed), and thus creating the largest cannabis-oriented pharmaceutical company in the world, by both revenue and market reach (Linnane, 2021). It was one of the companies that greatly benefited from legalizing recreational use of cannabis in specific U.S. states (namely New York, New Jersey, Montana, Virginia and others), with other company that falls in that same category, and that appears in the above table being SNDL.

Companies that are on the fourth spot have 3 appearances in the Table 4, namely TSLA, BB, NOK and BBBY. Tesla's popularity and frequent appearance in the media did not grant them many appearances on the lists of most popular stocks, which shows that either the makers of these lists decided to focus on companies that are newer and not as recognizable (which all of the stocks in the lists could be classified as, aside from Apple and AMD), or the methodology that they used preferred aspects that Tesla did not have. In addition, there are Blackberry (BB), Nokia (NOK) and Bed Bath & Beyond (BBBY), which all have been grouped together with AMC and GME during 2021, as their lesser-known counterparts that might experience a similar rise in stock price (Hasso, Müller, Pelster, & Warkulat, 2022). They attracted a lot of interest from online investors at the time since they had a substantial short interest, and a potential for a short squeeze that a lot of investors were anticipating. In case of BBBY this led to an increase in stock price from around $17 per share to $35.33 in the first two weeks of 2021. After January's increase all three of these stocks experienced a gradual decrease in price through the 2021 and early 2022, which would lead to their price, as of July 2022, being below or very close to that of Q4 2020.

Majority of the above table is composed of stocks that appeared on two or only one list of the selected publishers. Many of them have different reasons and justifications for being put on there, but this is where methodology of creating the lists comes into the forefront. Due to many different companies being present on just 10 lists, we can see that characteristics that a company should possess in order to gain some popularity on social media vary widely. Aside from companies that share the first four spots regarding how much they appear on the lists, which all were a part of some major event that took place, all of the other ones seem random in their positioning. Aside from previously mentioned stocks, on top of the lists also appear companies like CLNE, CLOV and EXPR – a renewable natural gas company, a healthcare and insurance provider, and a fashion retailer for young men and women. These companies share almost nothing in common, from their industry, business model, profitability or market share. Any similarity that does appear can largely be attributed to coincidence. With all this, a couple conclusions that loosely connect them can be established. Large number of these companies are focused on new and emerging technologies (TSLA, LCID, UPST, WKHS…), which is in line with type of retail investors on social media sites, young people that want to be early adopters of new technologies. Whether that be from an altruistic standpoint of wanting more environmentally friendly versions of things that already exist, or from a purely profit based reasoning of getting early into a company that has a lot of potential and may soon be at the top of its respective industry.

Second category can be classified as relatively older companies, like BB, AMC, GME, which were prominent in late 80s, through 90s and early 00s. Investors that are now participating in the discussion on social media, were born in this same period, so they have a connection with them, and are emotionally attached to the company itself. These companies have slowly been losing relevance and market share to others. Online investors appear to take this matter personally and will try to keep those companies relevant and operating by investing in them. The reasoning for why they do so lies in the field of behavioral finance, which is out of the scope of this research, but nonetheless appears in these examples, where investors are not entirely guided by profit, but by personal connection to certain businesses.

2.2 Five selected stocks

Resulting from analysis of popular stocks on social media in the previous chapter, we will select five different companies for which we will calculate expected stock price and compare it with the amount of social media discussion to showcase the correlation between the two. Those five companies/stocks are – Tesla (TSLA), GameStop (GME), AMC (AMC), Palantir (PLTR) and Nio (NIO).

Tesla has been one of the most discussed stocks on social media since its IPO in 2010, and it led the stock market into a new era of highly volatile and speculative stocks that

still managed to prosper in long term. It stands out due to its large market share in the electric vehicle industry, its enormous market capitalization which reached one trillion USD in 2021, significant presence on social media (both its CEO and the company itself), highly committed and loyal fanbase of both user and investors, and at same time large amounts of deniers, negative press, and shortfalls (Liu, 2021).

Second selected company is GameStop. As previously analyzed, it was one of the most discussed companies on the stock market in 2021 and a leading force for many of the events that happened that year, and as previously stated one of the initial reasons for writing this thesis. In addition, it is one of the most frequently appearing companies on the popularity lists mentioned above in Table 4.

Third company is AMC, which has been tightly connected to the events that happened with GameStop, and in media reports was usually grouped together with it. In the beginning of 2021, two companies shared a similar stock price movement pattern, but over the rest of the year, AMC separated itself, and experienced another rise in price that GME did not. This consequentially made its price have a higher percentage drop than GME which turned out to be more stable in the long term (see Figure 7). Calculations in this thesis will help show if AMC and GME had similar expectations in that time period, and if Twitter and Reddit had the same effect on both of them.

The fourth company for the research is Palantir, and it is the only one in research that appeared just once in Table 4. It stands out by having its IPO in 2020 and being relatively new on the stock market. In addition, it is heavily connected to the American government which also gives a unique perspective of a company that the U.S. administration might pay a much closer attention to (Pitchbook, 2022). Its lack of appearance in the aforementioned lists is due to it having its first massive rise in price in 2020, close to its IPO date. By middle of 2021, its stock price stabilized and reasons for discussion became less apparent.

The fifth company is Nio, a Chinese electric car manufacturer that saw a major increase in price during 2020. Its unique aspect is that it is the only non-American company on the list of our selected stocks, thus giving us ability to see if discussion on Twitter and Reddit, first of which is heavily restricted and the second of which is completely blocked by Chinese government (Truong, 2018), can influence a company that is not natively American and presumably out of its sphere of influence. Second reason for its inclusion is that it is a direct competitor to Tesla, and by comparing the results from TSLA and NIO we will be able to draw more conclusions on if they share similarities in this aspect or not.

2.3 Profitability analysis of five selected stocks

An analysis of profitability indicators for the five companies is conducted in order to analyze their financial position which will help in determining if the rise in stock price that these companies experienced was because of their business model and operations or due to the outside forces. Indicators used are – return on equity (ROE), return on assets (ROA), gross profit margin, net profit margin, operating profit margin, and earnings before interest, taxes, depreciation and amortization (EBITDA). For calculation of ROE a DuPont analysis was conducted, in which both Three-step DuPont, and Five-step DuPont equations were used.

Table 5: DuPont analysis for selected companies in 2020 and 2021

2020	TSLA (in millions USD)	GME (in millions USD)	AMC (in thousands USD)	PLTR (in thousands USD)	NIO (in thousands RMB)
Net Income	862	-215.3	239,979	-1,166,391	-5,304,082
Earnings Before Tax (EBT)	1,154	-269.9	402,379	-1,179,027	-5,297,714
Earnings Before Interest and Tax (EBIT)	1,994	-237.8	442,644	-1,173,679	-4,607,645
Revenue	31,536	5,089.80	2,814,956	1,092,673	16,257,933
Average Total Assets	52,148	2,472.60	5,246,338	2,690,504	54,641,929
Average Total Equity	22,225	436.7	643,101	1,522,550	27,170,956
Return on Equity (ROE)	3.88%	-49.30%	37.32%	-76.61%	-19.52%

2021	TSLA (in millions USD)	GME (in millions USD)	AMC (in thousands USD)	PLTR (in thousands USD)	NIO (in thousands RMB)
Net Income	5,644	-381.3	250,596	-520,379	-4,016,949
Earnings Before Tax (EBT)	6,343	-395.4	374,232	-488,494	-3,974,684
Earnings Before Interest and Tax (EBIT)	6,523	-368.5	489,922	-411,046	-4,496,303
Revenue	53,823	6,010.70	3,077,608	1,541,889	36,136,423
Average Total Assets	62,131	3,499.30	5,748,946	3,247,450	82,883,601
Average Total Equity	30,189	1,602.50	902,672	2,291,030	34,785,557
Return on Equity (ROE)	18.70%	-23.79%	27.76%	-22.71%	-11.55%

Source: Own work.

For calculation of the Three-step DuPont the following equation was used:

$$ROE = Net\ Profit\ Margin * Asset\ Turnover * Equity\ Multiplier \quad (1)$$

In the above equation *net profit margin* represents the net income divided by revenue of the selected company for specific year, and it is also used as one of the other profitability indicators in Table 6. *Asset turnover* is the measure of asset use efficiency and is calculated by dividing revenues with total assets. Lastly, the *equity multiplier* is the measure of financial leverage and computed by dividing assets with shareholder's equity. The five-step DuPont analysis is also calculated, in which the equation (1) is supplemented with two additional factors: *the tax burden* (net income / EBT) and *interest*

burden (EBT/EBIT), and the net profit margin being changed into *operating profit margin*.

$$ROE = Tax\ Burden * Interest\ Burden * Operating\ Income\ Margin * Asset\ Turnover * Equity\ Multiplier \quad (2)$$

The two methods do not show any difference in calculated return on equity, and the results obtained have been placed in the Table 6.

Table 6: Profitability indicators for selected stocks in 2020 and 2021

2020	ROE	ROA	Gross Profit Margin	Operating Profit Margin	Net Profit Margin	EBITDA
Tesla	3.88%	1.65%	21.0%	18.85%	2.73%	USD 4.3 bn
GameStop	-49.30%	-8.71%	24.74%	2.43%	-17.09%	USD -134 mn
AMC	37.32%	4.57%	-	15.72%	8.53%	USD 547 mn
Palantir	-76.60%	-43.35%	67.73%	-27.14%	-106.74%	USD -1.16 bn
Nio	-19.52%	-9.70%	11.52%	11.99%	-32.62%	RMB -3.56 bn

2021	ROE	ROA	Gross Profit Margin	Operating Profit Margin	Net Profit Margin	EBITDA
Tesla	18.70%	9.08%	25.30%	21.36%	10.49%	USD 9.4 bn
GameStop	-23.79%	-10.90%	22.42%	-7.23%	-28.29%	USD -304.1 mn
AMC	27.76%	4.36%	-	15.91%	8.14%	USD 584 mn
Palantir	-22.71%	-16.02%	77.98%	21.65%	-33.70%	USD -396 mn
Nio	-11.55%	-4.84%	18.87%	5.44%	-11.12%	RMB -2.79 bn

Source: Own work.

From the above indicators, it can be seen that companies differ widely from one another in terms of profitability. In both years, only ones that are being profitable are Tesla and AMC, with all of the indicators being positive.

According to the indicators, GameStop stands out with a significant improvement in ROE from year to year, but all of the other indicators marginally decreasing, with gross profit margin also decreasing by 2.23%, operating profit margin reaching negative percentage, net profit margin decreasing even further into negative percentage, and EBITDA falling by $170 mn. It should be taken into account that this decrease in performance happened in 2021, when GameStop was one of the most talked about companies on the stock market, if that was not the case, and the events of January 2021 did not happen (which were entirely out of GameStop's control), their position would likely be even worse. From this assessment, it can be concluded that the reasoning for GameStop's increase in stock price was not based on its financial performance or on their possible future development, since it is a company that is lagging behind on new trends in the industry and it is not able

to catch up to competitors, which in this case would be, Microsoft, Sony, and platforms like Steam and Epic Games.

Regarding its business model, AMC can also be put into category of companies that are lacking potential future development. However, an advantage that they have is that for them it is easier to make a shift from theaters to home media since they already own rights for showing specific movies and shows. Main difference between GME and AMC is that AMC is a company that is selling no actual physical products or merchandise, and this difference in business model is seen in finances as well. AMC does not report on cost of goods sold, so it was not possible to calculate its corresponding gross profit margin. As of YE2021 AMC is managing to stay profitable with net profit margin of 8.14%, largely stabile profitability indicators, with 4 out of 5 them staying within less than 1% change from 2020 to 2021, with the exception of ROE. As a company founded in the early 1900s, which already saw its peak in both popularity and profit, AMC is not expected to make any kind of large changes to their business model, and is currently in the late stage of its existence. Shift from theatres to home media was the unexpected change that disturbed AMC's business model, and it is currently in a stage where in order to compete with the likes of Netflix, HBO and others, it would require restructuring and a long-term oriented business plan. According to its finances and profitability indicators, this is something that they are not doing, and it proves us that AMC does not have a perspective future or any long-term plan to help them compete. Currently it is a stable company, but this does not warrant the sharp increases in their stock price that it has been going through in 2021.

Palantir's profitability is harder to adequately assess since it is a new company on the market, and Table 6 only show its first complete year and a three of months of public trading. Its ROE, ROA, EBITDA, operating profit margin and net profit margin have all been negative in 2020. In the year after, Palantir managed to turn the operating profit margin positive, notably increase its ROA and ROE, which is still negative as of YE2021, and improve its net profit margin from -106.74% to -33.70%. Even with these changes they are still not becoming profitable, but considering that only one full year has passed, the improvements on all indicators being significant, and the backing that Palantir has from the U.S. government as one of its clients, it has an ambitious long-term perspective. It is highly likely that if they continue in this direction, they will turn profitable in 2022 or 2023, and continue developing for the foreseeable future. From investors' perspective, Palantir would currently fall into a high risk, high reward category, because while its financial situation is currently weak, it has potential and backing to improve, while at the same time, if these proposed deals do not materialize, it is going to have a hard time attracting new clients and continuously develop. This gives investors that are risk takers, ability to be among early adopters of this new company and reap higher profits if it manages to accomplish its goals. Thus, its high returns on the stock market can be partially justified by its financial statements.

By comparing profitability indicators of Tesla and Nio, it can be seen that the industry of electric vehicles is not the sole reason behind Tesla's success. Nio as one if its biggest competitors in that field, is still trying to become profitable, while Tesla did so in 2020 and is rapidly progressing forward. From these indicators it can be seen that for Tesla all five indicators are positive and improving from year to year, which is not the case for other selected companies. Its EBITDA more than doubled during 2021, net profit margin increased from 2.73% to 10.49%, and its ROE increased from 3.88% to 18.70%, indicating that the company is becoming better at converting its equity financing into profit. The biggest drawback that Tesla has is its volatility, that in most cases is caused by the forces outside of its finances. As previously mentioned, Musk's behavior and actions are at least partially connected to the stock price of Tesla (Kim, Lee, & Suh, 2021). Additionally, being not only an electric vehicle company, but a tech company as well, increases the range of its competitors, some of which are solely focused of advancement of technology and it is a hard task for Tesla to adequately compete with them. Even when these other aspects are taken into account, Tesla shows that it can become profitable and continuously develop and sustain its business, which does partially justify rise in its stock price, but not to the extreme degree that was the case in the previous years.

If Nio is analyzed on its own, without any comparisons to Tesla, it can be seen that its finances are slowly improving, with 5 out of 6 profitability indicators having a positive change from 2020 to 2021. Most significantly net profit margin rising from -32.62% to -11.12%, and ROE going from -19.52% to -11.55%. This shows a positive trend, that will likely will lead into profitability in next couple of years. On the negative side, its stock price has been falling since early 2021, and while they are currently conducting expansion of their business, current economic climate and the company's inability to lower expenses are a difficult challenge that needs to be solved. Depending on how the expansion of its business ends, Nio has a high chance of turning profitable and being a reasonable competitor in the electric vehicle industry, with its own secured market share. For investors this can be seen as a good opportunity for an early entry into a company, but that does not justify sharp stock price increases that it had.

3 IMPACT OF SOCIAL MEDIA DISCUSSION ON SELECTED STOCKS

3.1 Methodology

To see if the movements of selected stock prices are connected to the discussion on Twitter and Reddit, an event study will be made. If the results of the event studies indicate that there is an outside force that pushed the stock prices higher than it was expected, assumption will be made that this force is the social media mentions, and this assumption will be tested by regressing social media mentions on the stock's returns, and by

calculating correlation between the two. If the results from the event study are positive, and the second regression indicates that there is a positive correlation between stock prices and mentions, it can be concluded that discussion on Twitter and Reddit does influence selected stocks to a certain degree.

The structure of the event study will follow the methodology published in the *Journal of Economic Literature* titled *Event Studies in Economics and Finance* (MacKinlay, 1997). In this paper, MacKinlay indicates that event studies measure the effect of a specific event on value of firms. The first step in this methodology is to specify the event window – a period of time during which the effect of an event will be analyzed. In general, the event window in this research will be formed around the event itself, which is decided to be the day on which stock price of a specific company experienced a sudden increase in price. By making that the event, the event window which will capture one year before and one year. This time frame was chosen since it allows to capture the movements of stock prices in the past and have that as a point of measure to see how much did it increase, and also to capture the movements after the event to see how long-lasting the impact was; if it was only momentary and started to disappear quickly, or if it persisted for a long period of time. This time frame of one year before and after the event will only be used as a template, since some of the stocks will require the window to be modified to better encompass the movements in price and happenings that occurred. In this study, the event window will be utilized as a period of time over which the expected return for every day will be calculated, for which the Fama & French 3 factor model will be used.

The Fama & French 3 Factor Model is an asset pricing model developed by Eugene F. Fama and Kenneth R. French in 1992 to describe stock returns (Fama & French, 1993). The model consists of three factors: the portfolio's excess return (portfolio's return minus the risk-free rate of return), SML (small minus big), and HML (high minus low). The SML factor indicates the outperformance of small versus big companies, and HML factor indicates the outperformance of companies with high price-to-book ratio versus the ones with low P/B ratio. This model represents an addition to the regular capital asset pricing model (CAPM) which uses only the first factor described above.

$$r = r_f + \beta_1(r_m - r_f) + \beta_2(SML) + \beta_3(HML) + \varepsilon \quad (3)$$

Fama French models will be calculated by taking the excess return of the specific stock for each day of the event window as the dependent variable and the daily values of the three factors, which will be obtained from the Kenneth French's website, as independent variables. The results of these regressions give us one coefficient for each of the three factors in the model. Those coefficients will then be multiplied by the same daily factors from French's website in order to obtain the daily expected returns. Expected daily returns are a fundamental part of being able to calculate the abnormal return. MacKinlay's paper indicates that in order to adequately appraise the impact of an event, the abnormal return needs to be measured, which in the paper are described as "…*the actual ex post return of*

the security over the event window minus the normal return of the firm over the event window. The normal return is defined as the expected return without conditioning on the event taking place" (MacKinlay, 1997).

$$AR_{i\tau} = R_{i\tau} - E(R_{i\tau}|X_\tau) \quad (4)$$

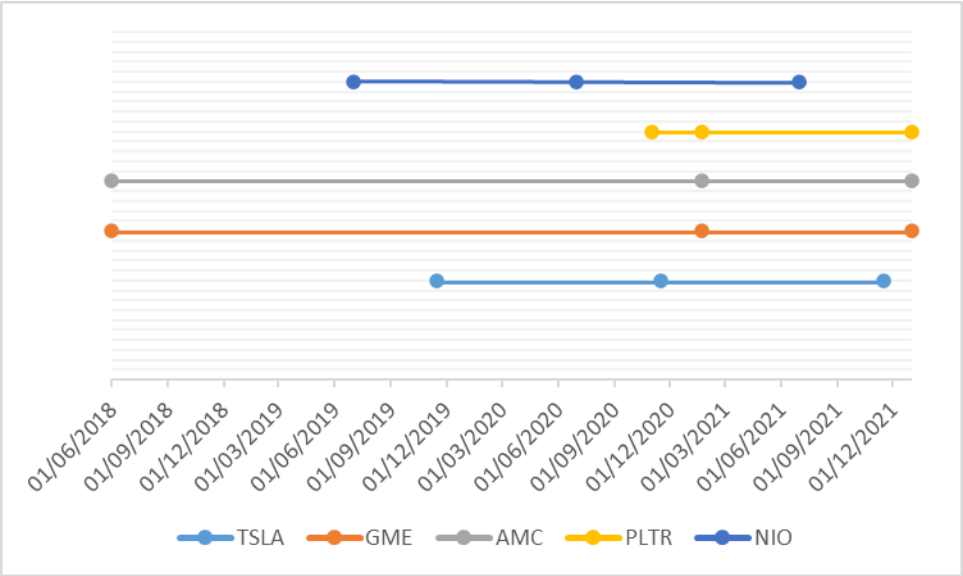
In formula (4), τ indicates the time frame, the $AR_{i\tau}$ is the abnormal return, $R_{i\tau}$ the actual return, and $E(R_{i\tau}|X_\tau)$ the normal return with the conditioning information, or in this case the expected return. In the case of this research, the abnormal return will help showcase how the stock compares to its expectations and if there is any possibility that social media discussion had any impact on the price. If for example, the AR over the event window for a stock is 100%, it indicates that some force is affecting the stock price, because stock outperforming its expectations by 100 percent in a specific time period is not a normal behavior. After obtaining ARs for all of the days of the event window they will be summed up in order to get the cumulative AR which gives the exact value of over/under performance during the event window.

After the results for AR are obtained, it will be determined how to proceed with the calculations, with three possible outcomes. First is that AR is going to be negative, which would indicate that the stock returns underperformed compared to its expectations. This is an outcome that this thesis is hoping to avoid since it would indicate that the chosen stocks have not been selected accordingly, and in the event window, there has been no actions or outside forces that led to an increase in stock price and its performance. Second outcome is that the AR is zero or a very low positive or negative percentage. This result would mean that the stock returns behaved exactly as expected or with small variations at points. This type of result is more in line with the thesis' expectations but still not adequate enough to prove the existence of events that influence the price. The third, and final possible outcome is that the AR is significantly positive. If this occurs, it has successfully been proven that the selected stock overperformed compared to its expectations.

That is where the second calculations of models take place. As this thesis studies correlation between daily amount of discussion on Twitter and Reddit and the stock price of five selected companies, and not causal relationships, correlations will be calculated and regressions will be conducted where daily stock returns will be put as a dependent variable, and daily mentions from Twitter and Reddit as independent variables. For each of the platforms a separate regression will be made. The specifics of data collection are described in the Chapter 3.2. After getting the results from regressions and correlations, we will conclude if the connection does or does not exist. If it does, and is positive, it could be concluded that a portion of abnormal returns could be contributed to social media mentions. And if it is zero or negative, it would mean that discussion on social media platforms does not affect the selected stocks.

As previously mentioned, for every of the five selected stocks, there will be an event specified on which the price of the company's stock started to experience large increase, and in accordance with it, the event window will also be constructed to encompass a large enough time period before and after. The original template for creating event windows is one year before and one year after the significant event, but in practice this proved to be difficult for many reasons. For example, in case of PLTR, whose IPO date was on 30/09/2020 which is very close to the selected event date of 12/01/2021. A notable date to mention is the 31st of December 2021, which at the date of calculating these models, was the last day for which the three Fama French daily factors were published, so it was chosen as an end date for three of the selected stocks, even if one year has not yet passed since the selected event. The figure below shows a visual representation of all of the selected dates, with the left dot representing the beginning of the event window, middle one the event itself and the right dot being the ending of the event window.

Figure 6: Event windows for selected stocks



Source: Own work.

As seen in the Figure 6, only two of the five stocks share the same window. The reasoning behind choosing the event date is primarily based on the analysis of the selected company's stock price, with the date being the day when a significant and notable percentage change happened, or the day which marked the beginning of a consecutive increases in stock price over a longer period.

For Tesla, the 16/11/2020 is taken as the event date. During this day Tesla experienced a sharp rise in stock price which would put the stock into a period of higher volatility for the next year. On that day the average closing price was $408.09, which is the last day on which Tesla's price remained around 400 dollars per share. During the course of the next day, the price increased to around $440, and the day after to $480. At the end of that week,

the closing price was $489, and it steadily increased to $600 in the next two weeks, and by the end of the year it passed the $700 per share mark. In the first week of 2021, it hit its peak of $880, which would remain the highest point for the price throughout most of the year, until November of 2021 when another sudden increase to around $1200 per share happened (for details on Tesla's stock price see Figure 7). Out of these two sudden spikes, we selected the first one, deeming it more significant in terms of percentage increase. The start and the end date of the event window are 15/11/2019 and 15/11/2021 respectively, marking exactly one year before and after the selected event, as indicated in the MacKinlay's previously mentioned research.

In case of GameStop, the selected event date is 21/01/2021. As described in Chapter 1.2.2., January 2021 has been a pivotal month for the company. By analyzing the stock price, January 21 was the beginning of a series of events that would lead GME's stock price to an all-time high. That week started with the price of the GameStop stock being at around $39, where it would mostly remain, but on 21st it increased to $43 per share, which can be seen as a start of the increase which would follow in the coming days. On 22nd the price reached $65, and after the weekend on January 25th it would reach a high of $159, with trading of the stock being halted nine times during the day (Lopatto, 2021). The stock hit its all-time high peak on January 28th which was $483, following which a sudden drop happened, with prices going below 100 dollars per share for the next couple of weeks, and then a gradual rise over the course of the year. Much of the same reasoning was used for selecting the event date as for Tesla, stock price analysis being the primary factor. Regarding the event window, as the ending point, 31/12/2021 was taken, since, it was the last day with available FF3 factors. The starting date had to be extended to 01/06/2018, for the following reason. GameStop as a company has been falling and deteriorating in the years prior to the events of January 2021, and with that being the main reason for its consequential increase, the research had to capture a much larger time span as opposed to only one year as indicated in MacKinlay's paper. If the starting date was to remain a year from the event, the starting stock price would be around $4.5, which is generally the price at which it remained until its eventual rise in late 2020, thus failing to capture the desired developments. By taking the date of 01/06/2018 we were able to capture its fall in stock price from $13 per share.

Regarding AMC, all three dates are the same as for GameStop. AMC was one of the companies that was heavily discussed during the events of January 2021, and it shares many similarities with GameStop, most notably for this research the stock price movement. From that, the reasoning behind the chosen dates remains the same as for GameStop. Event date of 21/01/2021 represents the last day on which the stock price remained in 'normal' range. With the following day, stock price increased from $2.98 to $3.5, and marked the beginning of its consequential rise. Additionally, the start of the event window is 01/06/2018 in order to capture the fall in price that the stock was experiencing in the past years, and 31/12/2021 is chosen as the end date for event window

with it being the last day with available FF3 factors at the time of calculation. Having the same reasoning and dates chosen as with GME, will clearly show the differences, between the two companies.

In choosing the event dates, Palantir was the most restrictive one. Its IPO date was on 30/09/2020, and in order to avoid the price volatility following the IPO, the first full month of public trading was skipped. The first date after was 02/11/2020, which was then taken as the starting date. The event date itself is similar to the previous two stocks, but it is one day after. On 22nd PLTR experienced its first significant increase in price in the two months. It increased from $25.98 which was the closing price on 21st, to a high of $32.65. In the day after it nearly reached $40 per share, and it remained in the high to mid-30s for the following three weeks, after which it returned to its price before the rise which was in the mid-20s. Since the time PLTR spent on the stock market was much different from GME and AMC, there was no direct need to take exactly the same date as the previous two companies for purpose of continuity, thus making its event date one day later, but much more adequate for its price movement. The end date of 31/12/2021 is again taken as the last day, due to availability of FF3 factors at the time of calculation. PLTR is the stock with significantly smaller event window than other selected companies, but considering it also spent significantly less time being publicly traded, its window does correctly represent the company's stock price development.

The last of the selected stocks is NIO. In addition to Tesla, it was the only other stock on which we could use the template of one year prior and after the event. In this case, the event date is 01/07/2020. Prior to it, the stock had a period of two months in which the price gradually increased from around 3 dollars per share, to a price of $7.91 on our chosen event date. Daily increases until this point have not been high enough to justify being selected as the event. On the chosen date, price increased to $7.91, then $9.38 on the following day, and to $11.51 on the day after. Following this, the price fluctuated around $15 per share in the following two months, after which it gradually increased to a high of around $60 by the end of the year. The event window for NIO contains this increase and also its gradual decline into around $50 per share in the first half of 2021. The start date of 01/07/2019 encompasses the rise in price from $2.6 of the first day of observation, to already mentioned event date. The start and end date (01/07/2021) of the window are in accordance with MacKinlay's paper, being exactly one year before, and after the event.

Stock price data was taken from the beginning of the event windows for the stocks until the latest available date, which at the time was 21/01/2022. Once the stock price data was obtained, it was modified to calculate the needed components for models. From the daily stock data, we primarily used the adjusted closing prices for the calculations, while the daily high and open prices were used to analyze the price on specific days to determine the best event date. Adjusted closing prices were used to calculated the daily return, and afterwards it was used to get the daily excess return which will then be applied as a

dependent variable in the Fama & French 3 Factor models. Visual representations of the movements in five selected stock prices, and the data obtained can be seen in the Figure 7.

Figure 7: Event window stock prices for selected stocks

Source: Own work.

TSLA and NIO have the event line positioned exactly in the middle, while others are located more in the left or right halves of the graph. By analyzing the graphs, three different types can be differentiated. First would be TSLA and NIO which show a gradual but still exceptionally high increase over the course of the event window, and afterwards a continuous rise in price. Second are GME and AMC, which show a stagnation and decrease in price until the event date, after which an extraordinary increase occurs, that is followed by an immediate drop in a time frame of a month. After the drop, the prices

stabilized at a new high. A difference can be seen between the two companies where AMC experienced another sharp rise a couple of months after, while GME started to lose its value. PLTR does not follow the same trends as the other ones, and after the event it experienced an increase, but proportionately much lower than the others, and in the long term it does not have the continual increase that TSLA and NIO do.

3.2 Analysis of data

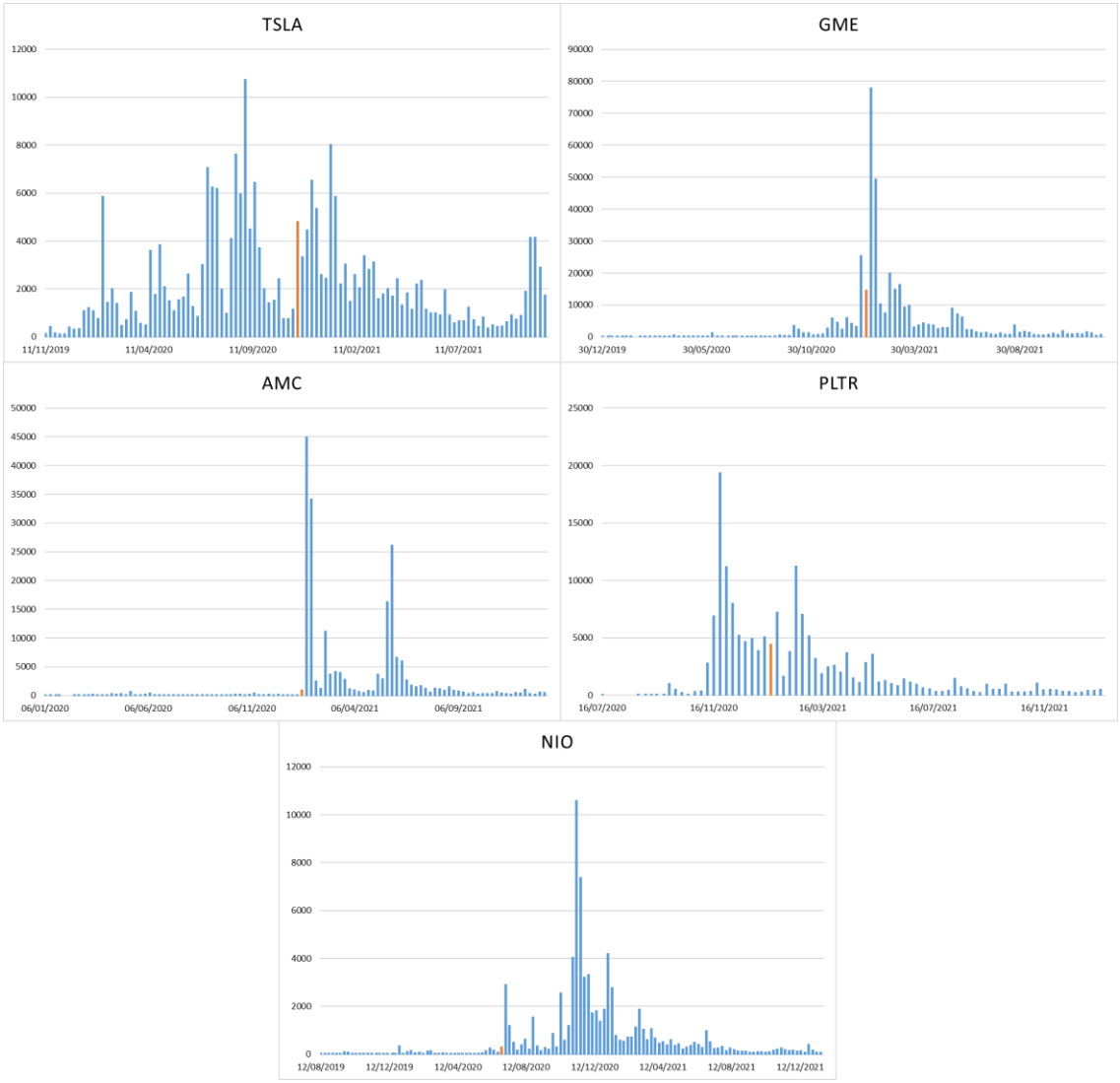
In order to properly calculate the Fama French models and mentions regressions, requiring data is the stock prices for selected stocks which were described in the previous chapter, and the following:

- Daily mentions for selected stocks from Reddit
- Daily mentions for selected stocks from Twitter

3.2.1 Reddit mentions

In order to see how much have the selected stocks been mentioned on Reddit, data from QuiverQuantitative database is used. It is a tool that, as described on their website "*helps retail investors to tap into the power of big data, and have access to actionable, easy to interpret data that hasn't already been dissected by Wall Street*" (QuiverQuantitative, 2021). It scrapes alternative stock data from the Internet, and compiles it on their on-line dashboard, through which daily stock mentions for Reddit are accessed. The availability ranges from 2018 until the current day, which falls in line with definitions of event windows. In Figure 8, weekly mentions on Reddit for each of the stocks are calculated and graphically showed. Weekly time frame is used instead of daily is to preserve the clarity of the graph by having the lower amount of data points for each stock. The orange line, that is approximately in the middle of each graph, is the week when the designated event of that stock took place. For the charts of GME, AMC and NIO, a portion of the beginning of the graph is cut in order to preserve the visual clarity of the chart and ensure easier data reading.

Figure 8: Weekly Reddit mentions in the Event Window for selected stocks



Source: Own work.

As seen in the figure above the stock prices and mentions on social media are not perfectly correlated. That can be seen most clearly in the examples of GME, AMC and NIO, where their stock price, after the selected event, settled on a new high; the stock mentions on the other hand, are mostly composed of spikes in specific weeks, and longer periods of lower activity. This indicates that social media participants will actively engage in conversations when the exceptional movements in stock prices are occurring, and when the price is stagnating, the engagement will be lower, since there is no news to be discussed.

If compared with the figure that represent the stock prices through the event windows, a couple notable characteristics can be pinpointed. Tesla has continuously been one of the most talked about stocks on the social media with its almost constant high mentions count. It has notable spikes but the surface area of the chart is much larger than that of the other ones. Also notable for Tesla, is that its highest point in mentions is barely above 10,000 in a week, compared to the GME and AMC, which reached 77,000 and 44,000

respectively. For both of those two stocks (and especially for AMC) in the week of the event, mentions are significantly lower than in the week after it. For PLTR, the initial spike that happened near the end of 2020 is predominantly contributed to the discussion regarding the IPO of the company, and we wanted to avoid the information and discussion that was focused on it by moving the event closer to that of GME and AMC. For PLTR we can also see that the stock price increase that happened immediately after the event did not induce an increase in mentions, but rather it happened with a lag of few weeks, which could mean that the discussion that occurred was predominantly of the negative sentiment and people were discussing the fall in price instead of its previous increase. NIO shares a lot of similarities with GME and AMC, with its lower discussion in the week of event compared to the week after it, and its one predominant spike in the middle of the event window.

If we look at the total number of mentions during the event, it can be seen which of the 5 selected stocks comparatively were the most popular ones. It will also show how significant the events themselves were, and amount of impact they had on stocks.

Table 7: Volume of mentions during the event window

	TSLA	*GME*	*AMC*	*PLTR*	*NIO*
Event Volume	233,228	376,358	200,747	156,678	73,264

Source: QuiverQuantitative (2021).

From the data gathered from QuiverQuantitative database, GameStop experienced the biggest amount of discussion on Reddit. Its number of mentions during the event window is 61% larger than the second placed TSLA with 233,228 mentions. This correlates with the excessive high that GME was experiencing during January 2021 with 77,000 mentions in a single week. TSLA and AMC appear to be in a similar range, even though their distributions of mentions through the event window is very different, as can be seen in Figure 8. In case of PLTR, its event volume of 156 thousand is significant considering its event window is much shorter than that of the other stocks. NIO has the lowest amount of mentions with less than 100 thousand. The reasoning for this might be the lack of a major public event, or a figurehead who would promote the company in the public media.

3.2.2 Twitter mentions

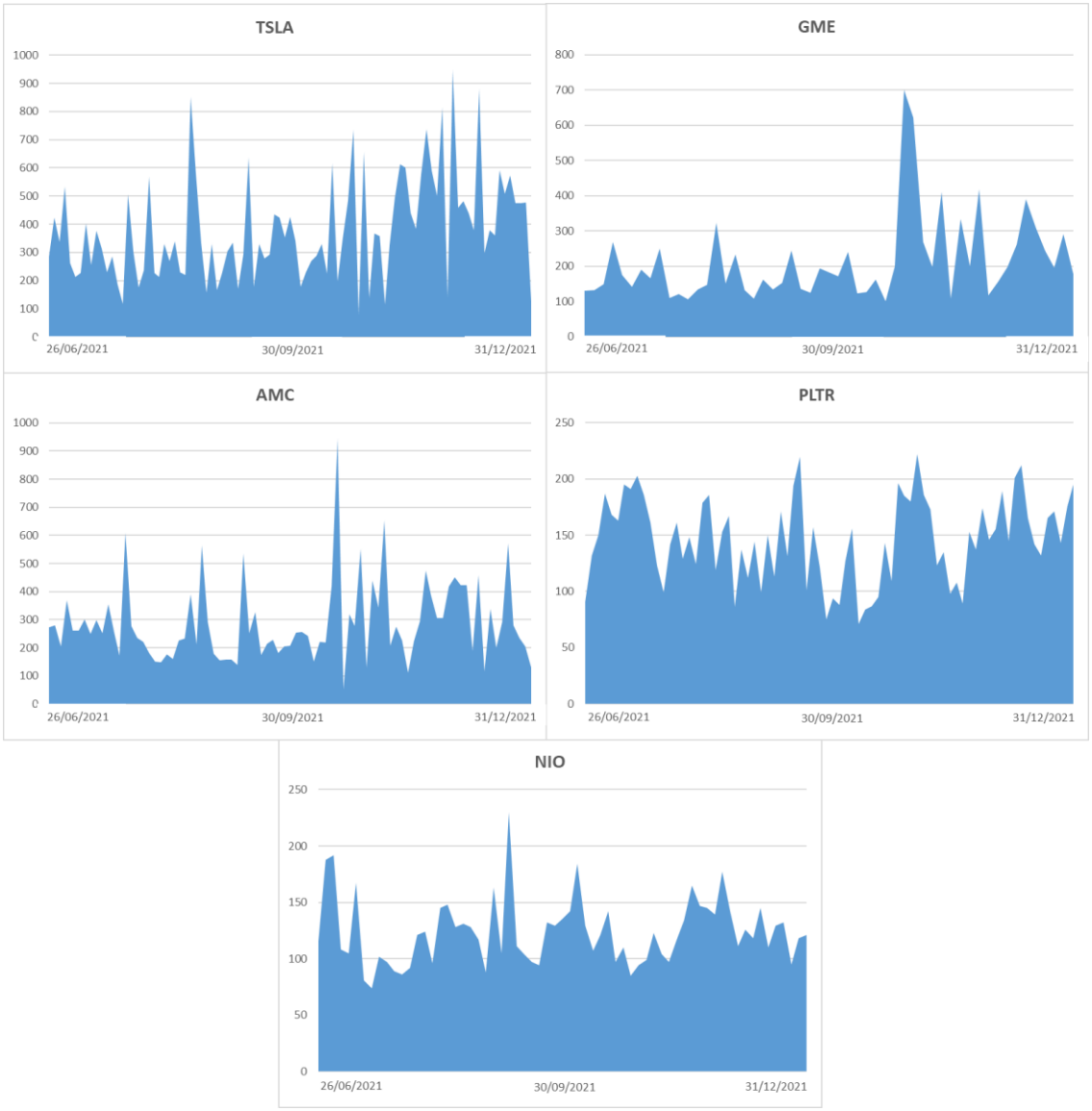
For Twitter mentions, due to the availability of data, we were able to gather data only for year 2021, from June until the end of the year. Due to this, the methodology of analysis of Twitter mentions will be different than those from Reddit. When the regressions are made of stock returns and Twitter mentions, they will analyze a smaller portion of the event window, as opposed to the whole one. This section will be in a period after the event, and will focus on the aftereffects. Meaning that after the event has played out and

all of its immediate effects have settled down, time frame of stock returns will be taken and analyzed with respect to Twitter mentions. The continuity of the research still holds, since before this point, the Fama French models will already be made for desired stocks and found out if there is an underlying effect, and while for Reddit mentions we tried to justify the whole event window, for Twitter we will try to justify only a portion.

The database used to gather information is the Tradytics tool/website. It is a tool that uses artificial intelligence and data analytics to help traders get more information about desired stocks (Tradytics, 2021). Most of its features require a paid monthly subscription, but in this case, a free version offers a detailed look on stock discussion on Twitter and measures daily how many times a certain stock ticker has been mentioned, percentage change in mentions since the previous day, average retweets, average likes and other. The website publishes data daily, which we compiled and separated according to the specific stock.

The figure below visually represents the most important data source from Twitter – daily mentions. Mentions are the primary information that we are going to use in further research and in calculation of models. When we compare them to the mentions on Reddit and data gathered in Figure 8, we can notice certain differences.

Figure 9: Daily Twitter mentions for selected stocks



Source: Own work.

Mainly there is a lack of significant, longer periods of increase or decrease in mentions, and the number of mentions is more consistent as opposed to Reddit ones. This can be partially attributed to the fact that there are shorter time periods for Twitter, but even when that is taken into account, in the example of TSLA, it was experiencing a constant, slow decrease in Reddit mentions in second half of 2021 and then a sudden increase in October/November, which in Twitter mentions can hardly be seen. From June to approximately beginning of November it was mostly constant between 200 and 400 mentions per day, with a couple of spikes on specific days. The rise above 400 started happening afterwards, but considering those spikes, and it previously being above 400 on a couple days a month, rise that eventually happened is hard to see on above figure. Data for GME and AMC overall looks similar to Reddit's. Second half of 2021, did not have any major increases or events happening and both Twitter and Reddit mentions are

stagnating over that period. Some similarities can be observed on the graphs themselves, that prove the tight connectedness of the two companies and their popularity – both of them reached their highest peak in mentions in Q4 2021, GME of exactly 700, and AMC of 944. Immediately after they dropped with a sharp decline. Chart for PLTR is similar to charts for its stock price and Reddit mentions, with this specific period being devoid of any major increases, only notable decrease is around September where the values dropped to around 50 per day. Lastly, the chart for NIO also appears to be very stable, with mentions mostly ranging between 100 and 200 per day.

3.3 Fama & French 3 factor models

In this section Fama & French 3 factor models will be calculated for the selected five stocks in order to assess by how much they over/under performed the expectations. We predict that stocks will have a noticeable overperformance compared to the expected return, which, if true, will be attributed to Twitter and Reddit discussion and consequently evaluated in Chapter 3.4.

Tables 8 to 12 contain results of the Fama French 3 factor models with its corresponding regression statistics and coefficients. A notable difference exists in the number of observations for five selected stocks, with one observation being one data point, or one day of the event window. GME and AMC have the highest number of 904, due to previously mentioned longer event window to capture earlier fall in price of the stock. TSLA and NIO have similar number of observations, 503 and 505 respectively. Both have event windows of exactly two years, and the difference of two days is due to two more non-working days or holidays being included in observations for TSLA since the two event windows started in different years.

Multiple R or the correlation coefficient shows how strong the linear relationship is between the variables. For all five of the selected stocks the values range between 0.26 and 0.57, which indicates a low to medium positive correlation. Two stocks with lowest correlation are GME and AMC, due to high number of observations and a discrepancy between stock prices of two companies and the values of Fama French factors, where one was constantly decreasing for a long period of time, while the other did not. This is further reinforced when observing the two stocks with highest correlation coefficient – TSLA and PLTR. Over their respective event windows, they have both kept a constant increase in stock price, without any significant falls, which mirrored the market in general at that time period.

The coefficient of determination (R squared) is used to explain how differences in one of the variables can be explained by differences in other variables. Five selected stocks from this research can be separated into two groups regarding this statistic. One with higher R squared – TSLA and PLTR, and the other with the lower – GME, AMC and NIO. The highest value of the R squared in the models is 32% by TSLA. Considering that in our

case, we are dealing not only with the mathematical returns of the stocks, but also with behavioral finance, which is the underlying factor in stock returns, it is difficult to confidently estimate it. When dealing with prediction of human behavior the coefficient of determination will tend to have values lower than 50% (Frost, 2018), because of the fact that the behavior that investors exhibit is not always following logical footpaths, and in many cases can be erratic and faulty. This is one of the explanations why we believe the values for R squared are in general low. Another possible explanation of this may be that the relationship between the variables is non-linear, so the model is not able to explain it. In addition to that, the three independent variables are statistically significant for all the models, which will be shown in the following chapter. This would point in a direction of models being more satisfactory and the issue of low R squared being less relevant. R squared is a type of statistic that works best when dealing with a sample of population (Henry, 2001), which in this case is not entirely true. While the event window is only a portion of the entire lifespan of a stock since its IPO to today, these models are not used to explain the entire lifespan, or population. Our case lies entirely within the event window and anything outside of that is beyond the scope of the research.

The standard error statistic describes by how much does the average of sample deviates from the average of the population. As described in the previous paragraph, in this case the average of sample does not exist since the event window is the entire dataset, and we are calculating the returns only within it, and the returns that stocks achieved before and after are not important for research. Due to this, standard error statistic is insignificant in this research.

Table 8: Fama French 3 factor model results for Tesla

Regression Statistics	
Multiple R	0.56966
R Squared	0.32451
Adjusted R Squared	0.32045
Standard Error	0.03826
Observations	503

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	3	0.35090	0.11697	79.90899	0.00000
Residual	499	0.73041	0.00146		
Total	502	1.08131			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	0.00308	0.00171	1.79952	0.07254	-0.00028	0.00644
Mkt-Rf	0.01461	0.00107	13.69768	0.00000	0.01252	0.01671
SMB	0.01036	0.00206	5.02797	0.00000	0.00631	0.01441
HML	-0.00879	0.00129	-6.82469	0.00000	-0.01132	-0.00626

Source: Own work.

For TSLA the first coefficient in the table is the intercept, or the alpha, which compares the returns with the benchmark. It is positive which means that there is excess return over

the benchmark, albeit a small one, but with the p-value of 0.0725 being more than 5% it is considered not statistically significant. For all three Fama French factors p-values are extremely small which results in statistical significance. Coefficients for Mkt-Rf and the SMB factors are positive and the HML factor is negative. This shows that TSLA in this selected period has more correlation with small stocks than the bigger ones. Considering the HML coefficient is negative, we can conclude that TSLA is more correlated with companies that have lower book-to-market values, also known as the growth stocks. These values are relatively low, with 0.0104 and -0.0088 respectively. Considering Tesla has a much higher market cap when compared to the other automobile companies (Root, 2021), its correlation with the small stocks provides evidence that it more closely resembles small stocks in its price movements, but with it being rather low, it should not be taken as identical movement patterns. HML coefficient being negative puts Tesla in the category of growth stocks, which are considered to have the potential to outperform the overall market over time due to their future developments. In addition, it has a very low book-to-market value of 0.04 as of March 2022, which also puts it into the category of growth stocks.

Table 9: Fama French 3 factor model results for GameStop

Regression Statistics	
Multiple R	0.28084
R Square	0.07887
Adjusted R Square	0.07580
Standard Error	0.09835
Observations	904

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	3	0.74531	0.24844	25.68678	0.00000
Residual	900	8.70464	0.00967		
Total	903	9.44996			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	0.00321	0.00328	0.97883	0.32793	-0.00322	0.00964
Mkt-Rf	0.00508	0.00241	2.11262	0.03491	0.00036	0.00980
SMB	0.03115	0.00457	6.81451	0.00000	0.02218	0.04012
HML	0.00986	0.00293	3.35964	0.00081	0.00410	0.01561

Source: Own work.

GME shares certain similarities with TSLA, namely the intercept being statistically insignificant while the 3 factors are statistically significant, and it having small values of calculated coefficients. Additionally, it has a positive relationship with small stocks, meaning that when small stocks are outperforming big ones this stock will increase in price. This indicates that GME is still connected to small stocks, regardless of its current high stock price. Two years before the events of January 2021 its stock price was in the single digit range, which made GME share movements with those stocks, and a lot of the correlation in the regression presumably comes from that period. The positive value of the HML factor indicates that it is more correlated with the value stocks. This correlation

is very small, and classifying GME as either value or growth stock would not be adequate, since its potential to outperform the market anytime in the future is unlikely as discussed in chapter 2.3. The company has been developing in a path of being a new digital marketplace, and revitalized itself after early 2021, but there are many people that believe it will eventually return to the place it has been at before, and its current financials are not indicating a prosperous long-term growth (Thomas & Repko, 2022).

Table 10: Fama French 3 factor model results for AMC

Regression Statistics	
Multiple R	0.25872
R Square	0.06694
Adjusted R Square	0.06383
Standard Error	0.12410
Observations	904

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	3	0.99433	0.33144	21.52207	0.00000
Residual	900	13.86019	0.01540		
Total	903	14.85452			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	0.00177	0.00414	0.42812	0.66867	-0.00635	0.00989
Mkt-Rf	0.00608	0.00304	2.00289	0.04549	0.00012	0.01204
SMB	0.03401	0.00577	5.89561	0.00000	0.02269	0.04533
HML	0.01321	0.00370	3.56950	0.00038	0.00595	0.02048

Source: Own work.

For AMC, the SMB has the highest coefficient; following that are the HML and Mkt-Rf factors. P-values are again below 0.05 for each of the three factors but above it for the intercept. In this case the explanations remain similar to the ones of GME. Correlation with smaller companies is reasonable considering the price at which it was trading before 2021, but in this case even more so since it did not reach the three-digit mark and stayed below $60 per share in its highest periods. Similarly, correlation with value stocks is showcased, but its value is again questionable for much of the same reasons as with GME.

Table 11: Fama French 3 factor model results for Palantir

Regression Statistics	
Multiple R	0.45938
R Square	0.21103
Adjusted R Square	0.20287
Standard Error	0.04416
Observations	294

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	3	0.15129	0.05043	25.85586	0.00000
Residual	290	0.56562	0.00195		
Total	293	0.71691			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	0.00225	0.00261	0.85857	0.39129	-0.00290	0.00739
Mkt-Rf	0.00746	0.00321	2.32131	0.02096	0.00113	0.01378
SMB	0.02104	0.00336	6.27023	0.00000	0.01443	0.02764
HML	-0.00772	0.00212	-3.63404	0.00033	-0.01190	-0.00354

Source: Own work.

Results for PLTR show that P-values for three factors are statistically significant and alpha is statistically insignificant. Results also indicate that PLTR has a small positive correlation to small stocks with a coefficient of 0.0210, and it also shows a coefficient of -0.0077 with the HML factor, which makes it more correlated with growth stocks.

Table 12: Fama French 3 factor model for NIO

Regression Statistics	
Multiple R	0.32469
R Square	0.10542
Adjusted R Square	0.10006
Standard Error	0.06234
Observations	505

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	3	0.22946	0.07649	19.67991	0.00000
Residual	501	1.94716	0.00389		
Total	504	2.17663			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	0.00380	0.00278	1.36431	0.17308	-0.00167	0.00927
Mkt-Rf	0.01031	0.00172	5.99494	0.00000	0.00693	0.01369
SMB	0.01475	0.00345	4.27154	0.00002	0.00797	0.02153
HML	-0.00446	0.00214	-2.08199	0.03785	-0.00868	-0.00025

Source: Own work.

NIO shows a positive correlation with small companies, and a negative correlation with the value companies, while like other selected companies has p-values below 5% for FF3

factors and above for the intercept. Nio again shows results akin to Tesla with the values of the coefficients being similar.

From Fama & French models we can conclude that TSLA, PLTR and NIO form one group that correlates more with growth stocks, while GME and AMC form the other with more correlation with value stocks, and all five of the selected stocks showcase a correlation with small companies on the stocks market. One of the results that stands out is significantly low value of the coefficients for all five stocks. This is due to the aforementioned volatility and the selected stocks having a price movement that is not usual in the rest of the companies that are included for the FF3 factors. Having low correlation with both small and big companies indicates once again that there is some outside influence, which other companies did not receive and indicates that 5 companies move in a unique way. As previously explained in Chapter 2.3, financial analysts have a hard time adequately valuating these companies, and classifying them as either growth or value stocks is much to the preference of the one doing the analysis. Low correlation in all of the coefficients enforces this thinking and further administrates their position on the stock market as unique cases.

3.3.1 Expected and abnormal return

After obtaining the coefficients from the Fama & French 3 factor models we use them to calculate the daily expected returns for each of the stocks in their respective event windows. We will subtract them from the actual returns in that specific period of time and get the abnormal returns, which will show by how much did stocks over/under perform expectations made by the FF3 model. Obtained results can be seen in the Table 13.

Table 13: Returns for the whole event window

	TSLA	*GME*	*AMC*	*PLTR*	*NIO*
Actual return	1338.78%	1108.44%	123%	72.77%	1857.69%
Expected return	352.64%	2318.60%	1959%	21.57%	656.99%
Abnormal return	226.50%	-52.14%	-90.64%	47.56%	173.38%

Source: Own work.

TSLA in its observation period of 503 days earned a cumulated return of 1338.78%, starting with a stock price of $70.43 per share and ending at $1013.39. During the same period, expected return calculated with the FF3 model was 352.64%. Difference between these two values equates to 226.5%. Resulting from this Tesla has significantly outperformed its expectations.

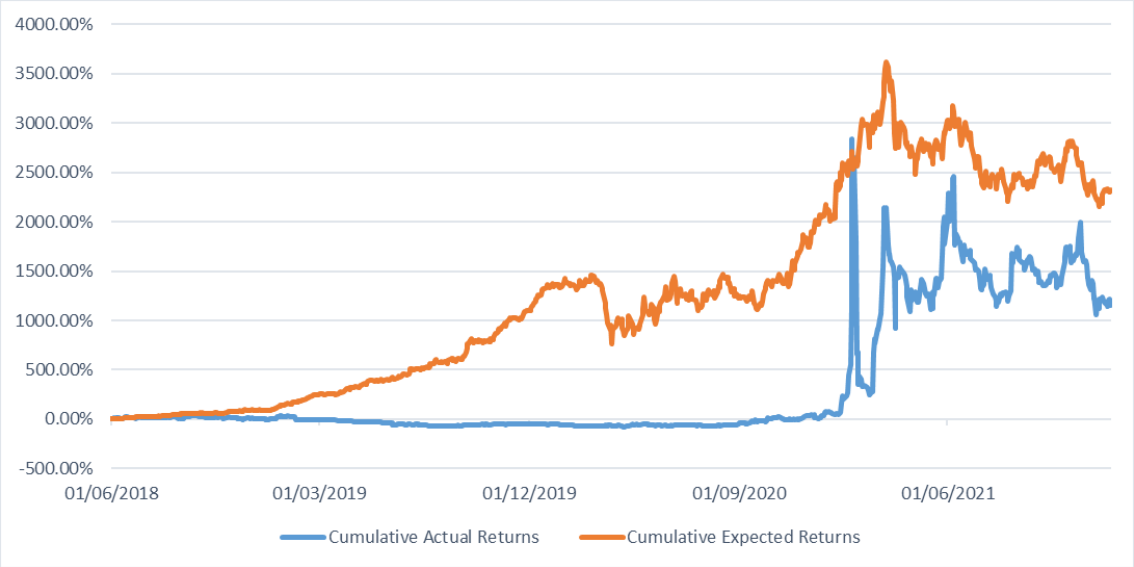
Figure 10: Cumulative actual and expected returns for TSLA during the event window

Source: Own work.

Presented in Figure 10 is a visual comparison of expected and actual returns for TSLA in the event window. In the first quarter of the event period its returns were in the expected area, which lasted until the end of June 2020. Afterwards it experienced the first of three major increases in price. This first increase lasted until 13th of November 2020, at which point cumulative returns have reached 484.66%, while the expected return at that point was 226.68%. Second increase happened immediately after with a sharp increase in the returns during the early 2021 followed by an immediate correction. The final wave started in mid-2021 where returns started to gradually increase without any sudden movements, although still at a high rate, followed by a sharp rise near the end of our observation period.

GameStop during its observation period of 904 days managed to realize a cumulative actual return of 1108.44% with a starting price of $12.28 per share and ending with $148.39. In the event period its expected return has been calculated to be at the level of 2318.6%, this leads to an abnormal return of -52.14%.

Figure 11: Cumulative actual and expected returns for GME during the event window

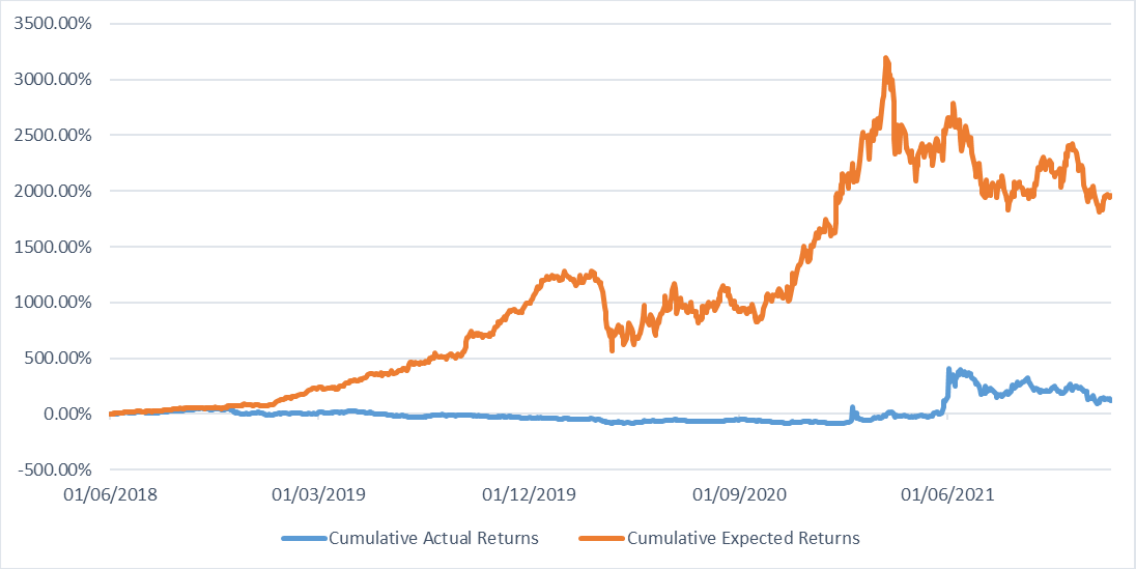


Source: Own work.

GME had a long period of negative returns which lasted until September 2020, afterwards the previously explained events took place and price experienced a sharp rise, which then stabilized at a new high. A major finding is that expected returns seem to be overestimated by the model. According to the model GME was expected to achieve approximately 1500% return during 2019, which knowing the state of the company at the time is severely exaggerated. Movements in the 2020 seem to be similarly overvalued. From cumulated return of approximately 1200% in the Q4 of 2020 to over 3500% in the first half of 2021. This kind of results seem to be unrealistic, and consequently we decided to modify the model to get a more accurate view. The model appears to produce high values of expected returns due to the uncommon stock price movements of GME, namely the long period of gradual decreases and then a sudden rise. In order to make the model more appropriate we decided to shorten the event window and take only the period from the event date until the end of the event window as a time frame. This will eliminate the aforementioned decrease in price and make the model produce a more realistic expectation. This alternative model will be constructed in the chapter 3.3.2, after which we will decide which one better simulates the expectations and more closely resembles the accurate picture.

AMC in its observation period of 904 days realized a cumulative actual return of 123%, with a starting price of $12.21 per share and ending at $27.2. In the event period, its expected cumulative return was 1959%, which leads to a difference of -90.64%.

Figure 12: Cumulative actual and expected returns for AMC during the event window

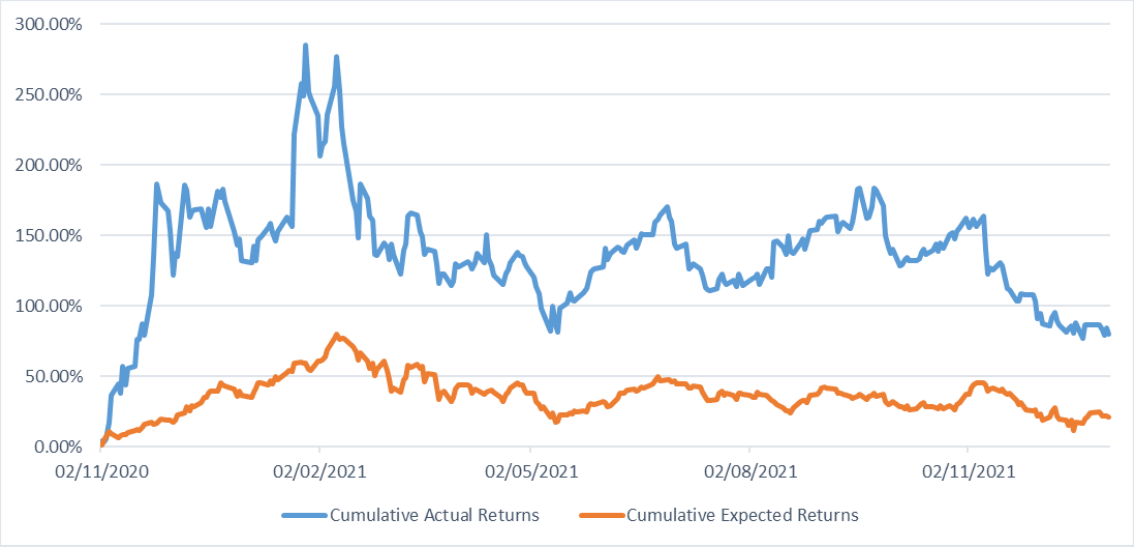


Source: Own work.

AMC according to this model has significantly underperformed the expectations. Similar to GME it has a long period of negative returns – from the end of 2018 until the mid-2021. In this time frame, its initial spike in stock prices, managed to get the cumulative returns above zero percent, but its consequential drop reduced it back below. It only managed to stay within positive returns after its second rise in price on June 2021. Second observation is that the expected returns line is essentially the same as the one for GME, only with marginally lower values. This also means that this model will suffer from the same overestimation of expected returns as that of GME, so we will use the same method of constructing a new model for AMC.

In case of PLTR, in its comparatively shorter observation period of 294 days, it has managed to realize a cumulative return of 72.77%, with a starting stock price of $10.54, and ending price of $18.21 per share. In that same event period, the cumulative expected return was 21.57%, which would indicate that difference between them, or the abnormal return is 47.56%.

Figure 13: Cumulative actual and expected returns for PLTR during the event window

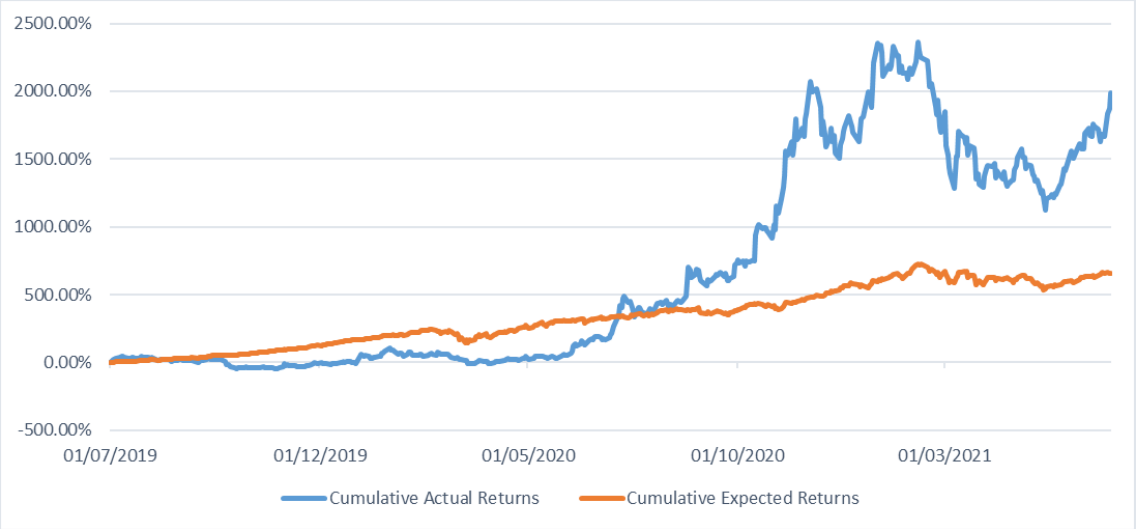


Source: Own work.

The figure above and the distribution of returns for PLTR is noticeably different from the previously analyzed stocks. By the end of 2020 actual returns immediately increased higher than the expected returns. Other than the initial spike at the start of the event window, the second most prominent characteristic is the second spike in the late January and early February of 2021. The actual returns reached their all-time peak at that point, which is in line with the expected returns which also noticeably increased albeit very gradually and without any sharp movements. This second spike also represents the selected event date. After it, the stock prices and consequentially the returns experienced a period of instability with three attempts at an increase but each resulted in a correction and return to the approximately same point as before. Expected returns were much less volatile during the same period, and the aforementioned spike in February 2021 was also its highest point, after which a slow and gradual decrease took place until the end of the observation period.

NIO in its observational period of 505 days managed to realize a return of 1857.69%, with a starting price of $2.6 and ending at $50.9 per share. For that same period the expected return calculated with the FF3 model was 656.99%, which makes a difference of 173.38% between the actual and expected cumulative return.

Figure 14: Cumulative actual and expected returns for NIO during the event window



Source: Own work.

The chart for NIO shares most similarities with that of TSLA, as was the case in previous aspects of this research. The biggest difference is the long period of negative abnormal returns, that lasted from the start of the event window, until the second half of 2020, and our selected event. In the days after the event there was enough momentum to turn the abnormal returns positive until June 2020. In Q4 of 2020 the price and consequently the returns started to exponentially increase, passing the 2000% for a brief moment but immediately falling below. Afterwards, one more rise happened, that was longer but also the response to it was a larger fall in returns, which only started to recover at the end of the observation period. Out of all other selected stocks, NIO had the most stabile expected returns line, with only a small decrease at the beginning of the Covid-19 crisis and a small rise at the start of 2021. Other than those, the line is on a constant upwards trend during the entire event window.

3.3.2 Alternative Fama & French models

Driven by the results we got for GME and AMC in the above models which resulted to be exaggerated and inaccurate, we will attempt to adapt the models to get an expected return line that is more realistic and in accordance with the selected company's characteristics. To achieve this, we will modify some of the research parameters. The following models will use the period from the selected event date until the end of the event window, shortening the observation period by removing the starting interval, where in case of GME and AMC the price and returns were drastically different. If we observe Figure 7 that showcases the stock prices, it can be seen that in these two cases the graphs are essentially split into two different parts, the beginning one, with a slowly falling stock price, and the subsequent one following the price increase. The following models will be made from data of the second part of the graph. By doing this we have eliminated a long

period of low prices and allowed the model to more accurately find the correct positioning for expected returns.

Table 14: Alternative Fama French 3 factor model for GME

Regression Statistics	
Multiple R	0.43751
R Square	0.19142
Adjusted R Square	0.18114
Standard Error	0.15273
Observations	240

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	3	1.30322	0.43441	18.62292	0.00000
Residual	236	5.50503	0.02333		
Total	239	6.80825			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	0.02215	0.00997	2.22211	0.02722	0.00251	0.04179
Mkt-RF	-0.04664	0.01239	-3.76332	0.00021	-0.07106	-0.02223
SMB	0.08877	0.01299	6.83627	0.00000	0.06319	0.11435
HML	0.01905	0.00903	2.10855	0.03604	0.00125	0.03684

Source: Own work.

Table 15: Alternative Fama French 3 factor model for AMC

Regression Statistics	
Multiple R	0.35623
R Square	0.12690
Adjusted R Square	0.11580
Standard Error	0.21331
Observations	240

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	3	1.56076	0.52025	11.43387	0.00000
Residual	236	10.73824	0.04550		
Total	239	12.29899			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	0.02983	0.01392	2.14255	0.03317	0.00240	0.05725
Mkt'RF	-0.06271	0.01731	-3.62278	0.00036	-0.09681	-0.02861
SMB	0.09768	0.01814	5.38625	0.00000	0.06195	0.13341
HML	0.01139	0.01261	0.90325	0.36731	-0.01346	0.03625

Source: Own work.

Tables 14 and 15 show the results of the regressions for GME and AMC, while the results for other stocks are presented in Appendix 3. By going through the new regression statistics and comparing them to the one of the original models, it can be seen that multiple

R statistic has improved for all of the stocks, with the smallest increase being in the case of TSLA, with an increase of only 0.0019. Other stocks have all experienced a rise in Multiple R in a range between 0.09 and 0.15, which, considering that the values of the statistic can range from 0 to 1 is a decent growth. This rise also indicates an improvement of the linear relationship in the models. The R Square statistic was improved as well, with TSLA having the lowest increase of 0.00216, GME and AMC, which were previously drastically low, have increased to 0.19 and 0.12 respectively, while PLTR and NIO saw further rise to 0.34 and 0.21. It can be concluded that shortening of observation periods in a way that has been described, improved statistical accuracy of the models.

Coefficients in the new models remained largely the same, with no changes from positive to negative and vice versa, only small corrections in the value of coefficients themselves. The p-values of the three Fama & French factors remain statistically significant for all five models, with the only exception being the Mkt-RF factor for PLTR.

Table 16: Returns for the alternative models

	TSLA	*GME*	*AMC*	*PLTR*	*NIO*
Actual return	148.33%	244.85%	812.75%	-29.91%	543.49%
Expected return	32.53%	-85.23%	-92.19%	-31.23%	79.37%
Abnormal return	87.09%	2030.59%	6345.30%	0.35%	267.32%

Source: Own work.

According to Table 16, some significant changes can be observed in the returns of selected stocks. Due to the shorter period of observation the cumulative returns are also much lower for all of the stocks except of AMC. This is attributed to a later date at which the alternative models started, thus making the difference between the beginning and ending price lower than in the initial models. Expected returns are also much lower, partially because of much shorter event window, and partially due to the lack of long period of time before the event with decreasing stock prices.

GME and AMC showcase a negative expected return over the selected period, that can be observed in more detail in the figures below (while figures for other stocks are presented in the Appendix 2).

Figure 15: Cumulated actual and expected returns for GME and AMC after event date

Source: Own work.

When compared to the graphs of the initial models, the difference is that actual returns are constantly above the line for expected returns. The line for actual realized returns is the same as in previous graphs, as the data remained unchanged. The expected return lines showcase a period of growth until March of 2021 which reflects the bull market that was happening at the time, but afterwards it is slowly decreasing until it reached the 0% return point and lower. This reflects the theoretical and financial expectations that were mentioned in chapter 2. As the alternative models better represent the conditions of GME and AMC, we have decided to take them for further research, and the initial models for TSLA, PLTR, and NIO.

3.4 Correlation between mentions and stock performance

In this chapter we are going to make a connection between the stock mentions and the returns that selected companies' stock managed to achieve in the specified period. Important note here is that we used the absolute return instead of the normal one. This is because a decrease in daily return, or it being negative, would mean a decrease and a lower number of mentions for the specific day if there was some correlation. This would be an inaccurate way of conducting the research since it would mean that if stock price sharply decreased on specific day, there would also be significantly less discussion about it on social media, while in fact it is quite the opposite. The discussion on social media would increase in both cases of sharp rise or a sharp fall in the stock price. Using absolute returns will help us circumvent this.

3.4.1 Tesla

Analysis will go through the companies one by one, with Tesla being first. Calculating firstly correlation between the variables.

Table 17: Correlations between mentions, returns and absolute returns for TSLA

	Reddit	***Returns***	***Abs. Returns***
Reddit	1		
Returns	0.1102	1	
Abs. Returns	0.4222	0.1960	1

	Twitter	***Returns***	***Abs. returns***
Twitter	1		
Returns	0.2846	1	
Abs. Returns	0.4021	0.1897	1

Source: Own work.

For correlation between Reddit mentions, daily returns and absolute daily returns, in case of Tesla, our assumption that mentions will behave more similar to absolute returns has proven to be correct. For both Reddit and Twitter, the correlation of 0.42 and 0.40 respectively, between daily mentions and absolute returns is noticeably higher than that of normal daily returns with mentions which stands at 0.11 and 0.28.

Table 18: Regression statistics and coefficients for TSLA

Regression Statistics	
Multiple R	0.40214
R Square	0.16172
Adjusted R Square	0.15250
Standard Error	0.01956
Observations	93

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	0.00672	0.00672	17.55497	0.00006
Residual	91	0.03482	0.00038		
Total	92	0.04153			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	0.00180	0.00468	0.38575	0.70058	-0.00748	0.01109
Twitter mentions	0.00005	0.00001	4.18987	0.00006	0.00002	0.00007

Regression Statistics	
Multiple R	0.42218
R Square	0.17824
Adjusted R Square	0.17666
Standard Error	0.03080
Observations	522

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	0.10697	0.10697	112.78688	0.00000
Residual	520	0.49317	0.00095		
Total	521	0.60014			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	0.01783	0.00183	9.75760	0.00000	0.01424	0.02142
Reddit mentions	0.00003	0.00000	10.62012	0.00000	0.00003	0.00004

Source: Own work.

In tables above are the results of regressions made with the data from Twitter and Reddit for TSLA. Multiple R (the absolute value of correlation between variables) is the same as correlation calculated in the Table 17. The R squared statistic falls between 17% and 16% for the two sites. In case where social media discussion is entirely responsible for the movements in stock prices the R squared statistic would be close of the value of 1. The values of 17.8% and 16.1% indicate that this much of stock's absolute returns are affected by daily discussion on the social media sites.

Regarding the coefficients of the regressions, both coefficients for Reddit and Twitter are statistically significant, with p-value under 5%. For Reddit we can conclude that if Reddit mentions increase by 1 on a specific day, then, ceteris paribus, correlation with absolute returns will increase by 0.00003. For Twitter, we can conclude that if Twitter mentions increase by 1, then, ceteris paribus, correlation with absolutes return will increase by

0.00005. If we consider that mentions are always counted in hundreds and in some cases in thousands, this equates to enough movement to explain the 17% and 16% of movement in absolute returns calculated through the R square statistic.

The intercept equals to 1.78% and 0.18% in Reddit and Twitter respectively, which would be the daily absolute return if there were no mentions on the specific day. Considering that dataset we have does not have any days when the mentions for either site equal zero, this coefficient is irrelevant. This will carry on further into the research for the remaining 4 stocks, since for all of them there is not a single day where mentions for either of social media platforms is zero.

3.4.2 GameStop

For GameStop absolute returns once again prove to have higher correlation than the normal ones when it comes to correlation of Reddit mentions.

Table 19: Correlations between mentions, returns and absolute returns for GME

	Reddit	***Returns***	***Abs. Returns***
Reddit	1		
Returns	0.2623	1	
Abs. Returns	0.7347	0.6395	1

	Twitter	***Returns***	***Abs. Returns***
Twitter	1		
Returns	0.1873	1	
Abs. Returns	-0.0558	-0.0010	1

Source: Own work.

The absolute returns exhibit a correlation of 0.73 with Reddit mentions, which is remarkably high. This gives evidence that the reason for the GameStop's movement since the beginning of 2021 is the discussion on Reddit. On the other hand, we can see that Twitter mentions have a negative correlation of -0.0558 with the absolute returns. This shows that, in the example of GME, Twitter did not play much of a pivotal role. Twitter mentions have a different time scope than those of Reddit so they are not including the early 2021, as explained in the previous chapter, but regardless, a clear distinction can be seen between the two platforms in their connectedness with GME.

Table 20: Regression statistics and coefficients for GME

Regression Statistics	
Multiple R	0.05580
R Square	0.00311
Adjusted R Square	-0.01643
Standard Error	0.01925
Observations	53

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	5.900E-05	5.900E-05	1.593E-01	6.915E-01
Residual	51	1.889E-02	3.704E-04		
Total	52	1.895E-02			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	0.02365	0.00540	4.38088	0.00006	0.01281	0.03449
Twitter mentions	-0.00001	0.00002	-0.39912	0.69147	-0.00005	0.00004

Regression Statistics	
Multiple R	0.73466
R Square	0.53973
Adjusted R Square	0.53785
Standard Error	0.10280
Observations	247

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	3.03639	3.03639	287.29649	0.00000
Residual	245	2.58936	0.01057		
Total	246	5.62575			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	0.02594	0.00706	3.67106	0.00030	0.01202	0.03985
Reddit mentions	0.00004	0.00000	16.94982	0.00000	0.00004	0.00005

Source: Own work.

The regressions made are further evidence of the correlation results. The R squared statistic is 0.54 for the Reddit model but only 0.003 for the Twitter one, with adjusted R squared for Twitter falling below zero. High value of R squared for Reddit indicates that more than half of the movement in absolute daily returns of GME is affected by daily discussion on Reddit. Regarding the coefficients of regression, for Reddit we can conclude that if mentions increase by 1 on a specific day, then, ceteris paribus, correlation with absolute returns will increase by 0.00004. The Reddit coefficient is also statistically significant with p-value lower than 5%. For Twitter mentions the p-value is above 0.05 and thus insignificant.

3.4.3 AMC

In case of AMC, correlations with Reddit mentions are similar to the results of ones for GME, with absolute returns showing much higher degree of connectedness than the normal ones with 0.72 and 0.52 respectively.

Table 21: Correlations between mentions, returns and absolute returns for AMC

	Reddit	***Returns***	***Abs. Returns***
Reddit	1		
Returns	0.5180	1	
Abs. Returns	0.7180	0.8634	1

	Twitter	***Return***	***Abs. Returns***
Twitter	1		
Returns	0.2521	1	
Abs. Returns	0.2351	0.2889	1

Source: Own work.

Twitter mentions correlation is very similar for both types of returns. In this case we have still decided to use the absolute ones for the regression, in order to be consistent in the research and since they better showcase the purpose of the research. Here we once again see that discussion on Reddit was much more impactful on AMC than discussion on Twitter, but at the same time Twitter correlation is much higher than that of GME which shows higher exposure that AMC has to Twitter, that GME is lacking.

Table 22: Regression statistics and coefficients for AMC and Twitter mentions

Regression Statistics	
Multiple R	0.23509
R Square	0.05527
Adjusted R Square	0.04360
Standard Error	0.04161
Observations	83

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	0.00821	0.00821	4.73853	0.03241
Residual	81	0.14027	0.00173		
Total	82	0.14848			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	0.02726	0.01027	2.65343	0.00959	0.00682	0.04770
Twitter mentions	0.00007	0.00003	2.17682	0.03241	0.00001	0.00013

Source: Own work.

Table 23: Regression statistics and coefficients for AMC and Reddit mentions

Regression Statistics	
Multiple R	0.71797
R Square	0.51548
Adjusted R Square	0.51350
Standard Error	0.14659
Observations	247

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	5.60105	5.60105	260.65438	0.00000
Residual	245	5.26467	0.02149		
Total	246	10.86572			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	0.02696	0.00988	2.73025	0.00679	0.00751	0.04641
Reddit mentions	0.00007	0.00000	16.14479	0.00000	0.00006	0.00008

Source: Own work.

In the regression statistics, the Multiple R statistic is the same as correlation between absolute returns and mentions. R squared shows a high value for Reddit and a low one for Twitter. R squared for Reddit indicates that 51.6% of all movements in absolute daily returns of AMC are affected by daily discussion on the Reddit, while for Twitter that is the case for only 5.5% of all movements in absolute daily returns. Coefficients have the same value for both platforms when rounded up to five decimal places – from that it results that if mentions increased by 1 in the specific day, then, ceteris paribus, correlation with absolute returns will increase by 0.00007. In both cases the p-values of the coefficients are lower than 5%, meaning that the coefficients are statistically significant.

3.4.4 Palantir

For PLTR, there is a large difference between correlation of Reddit mentions with normal returns and the absolute ones, being 0.20 and 0.57 respectively, with the value of absolute returns exhibiting a strong correlation.

Table 24: Correlations between mentions, returns and absolute returns for PLTR

	Reddit	***Returns***	***Abs. Returns***
Reddit	1		
Returns	0.1982	1	
Abs. Returns	0.5657	0.4249	1

	Twitter	***Returns***	***Abs. Returns***
Twitter	1		
Returns	-0.0510	1	
Abs. Returns	0.2261	-0.3620	1

Source: Own work.

In case of Twitter mentions the normal daily returns show a negative correlation of -0.05 while the absolute one is positive and with a moderate correlation of 0.23, which is lower

than that of Reddit mentions. This mirrors all the previous cases, in which it was also determined that Reddit discussion had more impact.

Table 25: Regression statistics and coefficients for PLTR

Regression Statistics	
Multiple R	0.22611
R Square	0.05113
Adjusted R Square	0.03830
Standard Error	0.01891
Observations	76

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	0.00143	0.00143	3.98715	0.04953
Residual	74	0.02646	0.00036		
Total	75	0.02789			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	0.00725	0.00874	0.83031	0.40904	-0.01015	0.02466
Twitter mentions	0.00012	0.00006	1.99679	0.04953	0.00000	0.00023

Regression Statistics	
Multiple R	0.56574
R Square	0.32007
Adjusted R Square	0.31782
Standard Error	0.02973
Observations	305

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	0.12608	0.12608	142.63086	0.00000
Residual	303	0.26785	0.00088		
Total	304	0.39393			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	0.01918	0.00204	9.38033	0.00000	0.01515	0.02320
Reddit mentions	0.00003	0.00000	11.94282	0.00000	0.00002	0.00003

Source: Own work.

The R squared statistic, albeit being much lower than in previous two cases, equals 32% for Reddit which indicates how much of stock's absolute returns are affected by daily discussion on the social media site. Comparatively, for Twitter the R squared statistic is only 5%. Coefficients for both Reddit and Twitter are statistically significant with p-value under 5%. For Reddit we can conclude that if mentions increase by 1 on a specific day, then, ceteris paribus, correlation with absolute returns will increase by 0.00003, while for Twitter the increase would be 0.00012.

3.4.5 Nio

The results for NIO stand out when compared to the other four selected stocks. It shows a very small difference between correlations of Reddit mentions and two types of returns, being 0.11 for 'normal' and 0.15 for absolute.

Table 26: Correlations between mentions, returns and absolute returns for NIO

	Reddit	***Returns***	***Abs. Returns***
Reddit	1		
Returns	0.1101	1	
Abs. Returns	0.1460	0.4778	1

	Twitter	***Returns***	***Abs. Returns***
Twitter	1		
Returns	0.0550	1	
Abs. Returns	0.4537	-0.2170	1

Source: Own work.

More importantly, it shows a relatively high correlation between Twitter mentions and absolute returns, which was not the case for any other stock. Nio appears to be the only stock in the research to be significantly more correlated with discussion on Twitter than that on Reddit. With the correlation being 0.45 between mentions and the absolute daily returns.

Table 27: Regression statistics and coefficients for NIO and Twitter mentions

Regression Statistics	
Multiple R	0.45370
R Square	0.20584
Adjusted R Square	0.19324
Standard Error	0.01926
Observations	65

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	0.00606	0.00606	16.32952	0.00015
Residual	63	0.02337	0.00037		
Total	64	0.02942			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	-0.01321	0.01025	-1.28836	0.20233	-0.03370	0.00728
Twitter mentions	0.00033	0.00008	4.04098	0.00015	0.00016	0.00049

Source: Own work.

Table 28: Regression statistics and coefficients for NIO and Reddit mentions

Regression Statistics	
Multiple R	0.14600
R Square	0.02132
Adjusted R Square	0.01926
Standard Error	0.04781
Observations	477

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	0.02365	0.02365	10.34558	0.00139
Residual	475	1.08594	0.00229		
Total	476	1.10960			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	0.04352	0.00239	18.23249	0.00000	0.03883	0.04822
Reddit mentions	0.00002	0.00001	3.21646	0.00139	0.00001	0.00004

Source: Own work.

The previous statement is further confirmed by the regressions made. There is a significant difference between the R squared statistic for two social media platforms, where the one for Reddit equals 0.02 and the one for Twitter 0.21, indicating that Twitter discussion affects around 20% of movements in stock's absolute returns, while Reddit discussion only 2%. Regarding the coefficients, for Reddit we can conclude that if mentions increase by 1 on a specific day, then, ceteris paribus, correlation with absolute returns will increase by 0.00002, for Twitter the situation is different, with one increase in mentions leading to a 0.00033 increase in the absolute return for that specific day.

CONCLUSION

From the analysis of current market conditions, we have discovered that a big part of all social media users have investments into stocks market (approximately 35%), with LinkedIn and Reddit having the highest percentage of participants. By looking at the ten different lists of the most popular stocks on social media in 2021 we identified two different characteristics that stocks share which make them more popular. First one being that a large group of stocks is oriented towards new and emerging technologies, while the second group are older companies whose stock price has been falling and whose business model is outdated, but investors hold sentimental value towards them and want to see them prosper.

We have taken five different stocks for further analysis – TSLA, GME, AMC, PLTR and NIO. From analysis of their stock price data and the amount of their daily mentions of Twitter and Reddit we concluded that the stock prices over event windows do not share similarities and there are not particular patterns that need to be showcased for the stock to be discussed on Twitter and Reddit. The amount of mentions for the two social media

platforms show that in general, mentions on Reddit are more condensed in one smaller period when the price sharply increased, and on Twitter they are more dispersed over the whole event without larger standouts. Afterwards, we utilized Craig MacKinlay's 1997 methodology for structuring event studies, to analyze by how much did the actual stock returns outperform or underperform the expectations. We have made two different Fama & French 3 factor models for all five of the selected stocks, one using the whole event window, and the other one made after adjustments, that uses period from the selected event until the end of the event window. For TSLA, PLTR and NIO we have determined the initial models to be more accurate, while for GME and AMC, alternative models showed higher accuracy. The results from the models are the following:

- TSLA has outperformed its expectations by 226.50%
- GME has outperformed its expectations by 2030.59%
- AMC has outperformed its expectations by 6345.30%
- PLTR has outperformed its expectations by 47.56%
- NIO has outperformed its expectations by 173.38%

After determining that all five stocks outperformed its expectations, we have conducted the analysis to see if this overperformance can be attributed to the discussion on Reddit and Twitter. We have done so by calculating correlation and making regressions between daily mentions of stocks' symbol on social media and daily absolute returns that they managed to achieve, with the results being as follows:

- TSLA showed correlation of 0.42 and 0.40 with daily mentions on Reddit and Twitter respectively; 17.8% of its absolute returns are affected by discussion on Reddit, and 16.1% by discussion on Twitter
- GME showed correlation of 0.73 and -0.06 with daily mentions on Reddit and Twitter respectively; 54.0% of its absolute returns are affected by discussion on Reddit, and 0.31% by discussion on Twitter
- AMC showed correlation of 0.72 and 0.24 with daily mentions on Reddit and Twitter respectively; 51.6% of its absolute returns are affected by discussion of Reddit, and 5.5% by discussion on Twitter
- PLTR showed correlation of 0.57 and 0.23 with daily mentions on Reddit and Twitter respectively; 32.0% of its absolute returns are affected by discussion on Reddit, and 5.1% by discussion on Twitter
- NIO showed correlation of 0.15 and 0.45 with daily mentions on Reddit and Twitter respectively; 2.1% of its absolute returns are affected by discussion on Reddit, and 20.6% by discussion on Twitter

In this thesis we have studied correlations between chosen stock prices and mentions on social media, not causal relationships. From the above results we can conclude that TSLA, GME, AMC and PLTR show significant correlation with daily discussion on Reddit, while TSLA and NIO show significant correlation with daily discussion on Twitter.

REFERENCE LIST

1. Ali, S., Umar, Z., Gubareva, M., & Yousaf, I. (2021). A tale of company fundamentals vs sentiment driven pricing: The case of GameStop. *Journal of Behavioral and Experimental Finance, 30*, 34-55.
2. Amos, C. (2021). *5 Top Reddit Stocks To Watch This Week Apart From GME & AMC*. Retrieved July 8, 2022 from https://www.nasdaq.com/articles/5-top-reddit-stocks-to-watch-this-week-apart-from-gme-amc-2021-06-10
3. Azar, P. D., & Lo, A. W. (2016). The wisdom of Twitter crowds: Predicting stock market reactions to FOMC meetings via Twitter feeds. *The Journal of Portfolio Management, 42*(5), 123-134.
4. Baer, D. (2022). *The 31 biggest subreddits (2022 update)*. Retrieved July 15, 2022 from https://blog.oneupapp.io/biggest-subreddits/
5. Bevan, J. (2022). *Robinhood Revenue and Usage Statistics (2021)*. Retrieved July 1, 2022 from https://www.mobilemarketingreads.com/robinhood-revenue-and-usage-statistics-2020/
6. Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of computational science, 2*(1), 1-8.
7. CFA Institute. (2016). *PAYMENT FOR ORDER FLOW - Internalisation, Retail Trading, Trade-Through Protection, and Implications for Market Structure*. Retrieved June 17, 2022 from https://www.cfainstitute.org/-/media/documents/issue-brief/payment-for-order-flow.ashx
8. Chohan, U. W. (2021). Counter-hegemonic finance: The GameStop short squeeze. *Financial Crises eJournal. 2021*. 10-15.
9. Clor-Proell, S. M., Guggenmos, R. D., & Rennekamp, K. (2020). Mobile devices and investment news apps: The effects of information release, push notification, and the fear of missing out. *The Accounting Review, 95*(5), 95-115.
10. Curry, D. (2022a). *Finance App Revenue and Usage Statistics (2022)*. Retrieved June 22, 2022 from https://www.businessofapps.com/data/finance-app-market/
11. Curry, D. (2022b). *Stock Trading & Investing App Revenue and Usage Statistics (2021)*. Retrieved June 25, 2022 from https://www.businessofapps.com/data/stock-trading-app-market/
12. Curry, D. (2022c). *eToro Revenue and Usage Statistics (2022)*. Retrieved June 27, 2022 from https://www.businessofapps.com/data/etoro-statistics/
13. Dorfleitner, G., & Scheckenbach, I. (2021). Trading activity on social trading platforms–a behavioral approach. *Journal of Risk Finance. 23*(1), 32-54.
14. Drake, M. S., Thornock, J. R., & Twedt, B. J. (2017). The internet as an information intermediary. *Review of Accounting Studies, 22*(2), 543-576.
15. Duggan, W. (2022). *U.S. News. 8 Reddit Stocks Trending Now*. Retrieved July 8, 2022 from https://money.usnews.com/investing/stock-market-news/articles/top-reddit-stocks-gaining-buzz
16. Eichengreen, B. (2022). European Inflation in an American Mirror. *Intereconomics, 2022*(2), 76-78.
17. Elliott, W. B., Grant, S. M., & Hodge, F. D. (2018). Negative news and investor trust: The role of $Firm and #CEO Twitter use. *Journal of Accounting Research, 56*(5), 1483-1519.

18. Eurostat. (2022). *Annual net earnings*. Retrieved June 29, 2022 from https://ec.europa.eu/eurostat/databrowser/view/earn_nt_net/default/table?lang=en
19. Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics, 33*(1993), 3-56.
20. Fox, M. (2022). *Business Insider. Tesla and Lucid top 10 most talked about stocks on Reddit's WallStreetBets forum*. Retrieved July 8 2022, https://markets.businessinsider.com/news/stocks/top-10-wallstreetbets-reddit-stocks-tesla-lucid-most-mentioned-tsla-2022-3
21. Frost, J. (2018). *How High Does R-squared Need to Be?.* Retrieved July 1, 2022 from https://statisticsbyjim.com/regression/how-high-r-squared/
22. Gu, C., & Kurov, A. (2020). Informational role of social media: Evidence from Twitter sentiment. *Journal of Banking and Finance, 121*(2020), 174-193.
23. Hasso, T., Müller, D., Pelster, M., & Warkulat, S. (2022). Who participated in the GameStop frenzy? Evidence from brokerage accounts. *Finance Research Letters, 45*(2), 102140.
24. Henry, N. W. (2001). *R-square and Standardization in Regression*. Retrieved June 3, 2022 from http://www.people.vcu.edu/~nhenry/Rsq.htm
25. Jackson, A. (2022). *Inflation sits at a 40-year high. So you may want to rethink this 'rule' about how much of your income to save and invest*. Retrieved June 15, 2022 from https://www.marketwatch.com/picks/how-much-money-should-you-really-be-investing-during-periods-of-higher-inflation-01635875513#:~:text=Experts%20generally%20recommend%20setting%20aside,which%20we%20will%20discuss%20below
26. Kabir, U. (2021). *Yahoo Finance. 10 Reddit Stocks that Tripled in 2021*. Retrieved 8th July 2022 from https://www.yahoo.com/?err=404&err_url=https%3a%2f%2ffinance.yahoo.com%2fnews%2f10-reddit-stocks-tripled-2021-171456020.html
27. Kim, D. P. K., Lee, J., & Suh, J. (2021). Elon Musk's twitter and its correlation with Tesla's stock market. *International Journal of Data Science and Analysis, 7*(1), 13-19.
28. Kopestinsky, A. (2022). *What Is the Average American Income in 2022?*. Retrieved July 10, 2022 from https://policyadvice.net/insurance/insights/average-american-income/
29. Lambert, L. (2021). *Robinhood's brand is severely damaged: 56% of account holders are considering leaving the app*. Retrieved July 9, 2022 from https://fortune.com/2021/02/19/robinhood-brand-damage-gamestop-hedge-funds-wallstreetbets-reddit-stocks-gme-amc/
30. Liadze, I., Macchiarelli, C., Mortimer-Lee, P., & Juanino, P. S. (2022*).* The economic costs of the Russia-Ukraine conflict. *National Institute of Economic and Social Research Policy Paper, 32*, 1-10.
31. Liew, J. K. S., & Wang, G. Z. (2016). Twitter sentiment and IPO performance: A cross-sectional examination. *The Journal of Portfolio Management, 42*(4), 129-135.
32. Linnane, C. (2021). *Tilray and Aphria meeger closes and creates world's biggest weed company by revenue*. Retrieved June 8, 2022 from https://www.marketwatch.com/story/tilray-and-aphria-merger-closes-and-creates-worlds-biggest-weed-company-by-revenue-11620064759

33. Liu, S. (2021). Competition and valuation: a case study of Tesla Motors. *IOP Conference Series: Earth and Environmental Science. 692*(2), 1-10.
34. Long, C., Lucey, B. M., & Yarovaya, L. (2021). "I Just Like the Stock" versus" Fear and Loathing on Main Street": The Role of Reddit Sentiment in the GameStop Short Squeeze. *SSRN Electronic Journal, 31*(10), 176-201.
35. Lopatto, E. (2021). *How r/WallStreetBets game the stock of GameStop, The Reddit meme that bludgeoned a hedge fund*. Retrieved June 27, 2022 from https://www.theverge.com/22251427/reddit-gamestop-stock-short-wallstreetbets-robinhood-wall-street
36. MacKinlay, A. C. (1997). Event studies in economics and finance. *Journal of Economic Literature, 35*(1), 13-39.
37. Mackintosh, P. (2022). *Short Interest in Decline*. Retrieved May 28, 2022 from https://www.nasdaq.com/articles/short-interest-in-decline
38. MarketWatch. (2022). *Overview SPDR S&P 500 ETF Trust*. Retrieved July 25, 2022 from https://www.marketwatch.com/investing/fund/spy
39. Mercer, M., Kadous, K., & Zhou, Y. (2022). Why do investors rely on low-quality investment advice? Experimental evidence from social media platforms. *Behavioral and Experimental Finance eJournal, 21*, 135-180.
40. Morris, C. (2021). *Here are the 5 stocks Reddit's WallStreetBets are talking about today*. Retrieved July 8, 2022 from https://fortune.com/2021/06/11/amc-gme-stock-wallstreetbets-amc-gamestop-meme-stocks-stock-picks-reddit/
41. Nasli, A. (2022). *Brokerage account sizes - Research by BrokerChooser*. Retrieved June 3, 2022 from https://brokerchooser.com/education/news/data-dashboard/brokerage-account-sizes#totalassetsinQ32021perbroker
42. PitchBook. (2022). *Palantir Technologies General Information and Overview*. Retrieved May 30, 2022 from https://pitchbook.com/profiles/company/43117-84
43. Rennekamp, K. M., & Witz, P. D. (2021). Linguistic Formality and Audience Engagement: Investors' Reactions to Characteristics of Social Media Disclosures. *Contemporary Accounting Research, 38*(3), 1748–1781.
44. Rodriguez, V. (2021). *CNBC|Momentive Poll: "Invest in You" August 2021. CNBC & Momentive*. Retrieved July 10, 2022 from https://www.surveymonkey.com/curiosity/cnbc-invest-in-you-august-2021/
45. Root, A. (2021). *Tesla's Market Cap Hit $1 Trillion. Here's the Math*. Retrieved May 28, 2022 from https://www.barrons.com/articles/tesla-market-cap-trillion-dollars-51635189306
46. Rothman, T. (2019). Trading the Dream: Does Social Media Affect Investors Activity - The Story of Twitter, Telegram and Reddit. *International Journal of Financial Research, 10*(2), 147-152.
47. Securities and Exchange Commission (SEC). (2020a). *Investor Alert: Frauds Targeting Main Street Investors*. Retrieved July 4, 2022 from https://www.investor.gov/introductioninvesting/general-resources/news-alerts/alerts-bulletins/investor-alerts/frauds
48. Securities and Exchange Commission (SEC). (2020b). *Investor Alert: Look Out for Coronavirus Related Investment Scams*. Retrieved July 4, 2022 from https://www.sec.gov/oiea/investoralerts-and-bulletins/ia_coronavirus
49. Securities and Exchange Commission (SEC). (2021). *Investor Alert: Understand the Significant Risks of Short-Term Trading Based on social media*. Retrieved July

4, 2022 from https://www.sec.gov/oiea/investor-alerts-and-bulletins/risks-short-term-trading-basedsocial-media-investor-alert
50. Tandon, C., Revankar, S., & Parihar, S. S. (2021). How can we predict the impact of the social media messages on the value of cryptocurrency? Insights from big data analytics. *International Journal of Information Management Data Insights, 1*(2), 100035.
51. Tenebruso, J. (2021). *Why Sundial, AMC, Blackberry, and Other Reddit Stocks Soared Today*. Retrieved July 8, 2022 from https://www.fool.com/investing/2021/02/24/why-sundial-amc-blackberry-and-other-reddit-stocks/
52. Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance, 62*(3), 1139-1168.
53. Thomas, L., & Repko, M. (2022). *GameStop launches NFT marketplace as it hunts for growth in digital-fueled turnaround*. Retrieved June 21, 2022 from https://www.cnbc.com/2022/07/11/gamestop-launches-nft-marketplace-as-it-hunts-for-growth-online.html
54. Tradytics. (2021). *The most talked about stocks at twitter on daily and weekly basis*. Retrieved June 3, 2022 from https://tradytics.com/twitter-chatter
55. Truong, A. (2018). *Users in China found Reddit blocked over the weekend*. Retrieved May 27, 2022 from https://qz.com/1354441/china-blocked-reddit-but-access-is-returning-for-some-users/
56. Van Doorn, P. (2021). *We put AMC, GameStop and other meme stocks' number to the test – here's which ones came out on top*. Retrieved July 8, 2022 from https://www.marketwatch.com/story/we-put-these-eight-meme-stocks-through-a-rugged-analytical-test-which-are-poised-for-growth-and-which-have-big-downsides-11622810160
57. Zaremba, A., Umar, Z., & Yousaf, I. (2021). Co-movements between Heavily Shorted Stocks during a Market Squeeze: Lessons from the GameStop Trading Frenzy. *Research in International Business and Finance, 58*(1), 1-9.
58. YouGov. (2021). *Share of social media users in the United States who have money invested in the stock market as of January 2021*. Retreived June 25, 2022 from https://www.statista.com/statistics/1255319/us-social-media-users-stock-investments-2021/
59. Warner, J. (2021a). *Top Reddit stocks to watch*. Retrieved July 8, 2022 from https://www.cityindex.com/en-uk/news-and-analysis/top-reddit-stocks-to-watch/
60. Warner, J. (2021b). *Reddit stocks: what stocks are trending?*. Retreved July 8, 2022 from https://www.forex.com/en/market-analysis/latest-research/top-reddit-stocks-to-watch/
61. White, W. (2021). *The 10 Most Talked About Stocks on Reddit Today*. Retrieved July 8, 2022 from https://investorplace.com/2021/06/the-10-most-talked-about-stocks-on-reddit-today/
62. QuiverQuantitative. (2021). *r/wsb Mention Trends*. Retreived June 10, 2022 from https://www.quiverquant.com/sources/wallstreetbets?utm_source=hs_email&utm_medium=email&_hsenc=p2ANqtz-9DLYi1UqW8C7qMZfyi1f93aa8SH_zT9DrQcK4LiMz3vW9xt_naP-x7z8vcl9NyiZBkagRk

APPENDICES

Appendix 1: Povzetek (Summary in Slovene language)

Z analizo trenutnih razmer na trgu smo ugotovili, da ima velik del vseh uporabnikov družbenih omrežij naložbe na delniških trgih (približno 35 %), pri čemer imajo uporabniki omrežij LinkedIn in Reddit največji odstotek udeležencev. Eden od glavnih razlogov za to so razmeroma nove spletne trgovalne platforme brez provizij, katerih število uporabnikov nenehno narašča, največje povečanje pa je bilo v letih 2020 in 2021. Povprečni uporabnik teh platform opravi transakcijo v vrednosti le 270 USD, kar je bistveno manj kot pri drugih trgovalnih platformah. Ti uporabniki imajo povprečne letne plače, tako v ZDA kot v Evropi. Iz drugih izvedenih raziskav lahko sklepamo, da vlagatelji, ki so začeli investirati po letu 2020, te spletne platforme uporabljajo predvsem za sklepanje poslov, informacije glede naložb črpajo predvsem iz družbenih medijev, velik obseg naložb pa imajo v posameznih delnicah in kriptovalutah. Vlagatelji, ki so začeli poslovati pred letom 2020, uporabljajo borznega posrednika ali agenta za vodenje svojih poslov, naložbene nasvete in napotke pa dobivajo od istih agentov ali s spletnih strani, ki se ukvarjajo s financami in naložbami. Le-ti imajo tudi veliko bolj stabilen in raznolik portfelj in določeno mero averzije do kriptovalut, zaradi česar trdimo, da obstaja jasna razlika med obema skupinama in da ima nova generacija vlagateljev drugačne značilnosti, motivacijo in cilje.

S preučitvijo desetih različnih seznamov najbolj priljubljenih delnic na družbenih omrežjih v letu 2021 smo identificirali dve različni značilnosti, ki sta skupni delnicam, zaradi česar so te bolj priljubljene. Prva je ta, da je velika skupina delnic usmerjena v nove in nastajajoče tehnologije, medtem ko drugo skupino sestavljajo starejša podjetja, katerih cena delnic pada in katerih poslovni model je zastarel, vendar imajo vlagatelji do njih sentimentalno vrednost in si želijo, da bi uspele.

V nadaljevanju smo analizirali pet različnih delnic – TSLA, GME, AMC, PLTR in NIO. Z analizo njihovih podatkov o cenah delnic in količine njihovih dnevnih omemb na Twitterju in Redditu smo ugotovili, da tečaji delnic v časovnih oknih dogodkov niso podobni in da ni posebnih vzorcev, ki bi jih bilo treba prikazati, da bi se o delnicah razpravljalo na družbenih omrežjih Twitter in Reddit. Količina omemb za obe platformi družbenih medijev kaže, da so na splošno omembe na Redditu bolj zgoščene v enem krajšem obdobju, ko se je cena močno povečala, na Twitterju pa so bolj razpršene skozi celoten dogodek, ne da bi prihajalo do večjih izstopanj. Nato smo uporabili metodologijo Craiga MacKinlayja iz leta 1997 za strukturiranje študij dogodkov, da bi analizirali, za koliko so dejanski donosi delnic presegli ali zaostali za pričakovanji. Za vseh pet izbranih delnic smo izdelali dva različna 3-faktorska modela Fama in French; enega, ki uporablja celotno okno dogodkov, in drugega, narejenega po prilagoditvah, ki uporablja obdobje od izbranega dogodka do konca okna dogodka. Za delnice TSLA, PLTR in NIO smo ugotovili, da so začetni modeli natančnejši, medtem ko so za delnici GME in AMC alternativni modeli pokazali večjo natančnost. Rezultati modelov so naslednji:

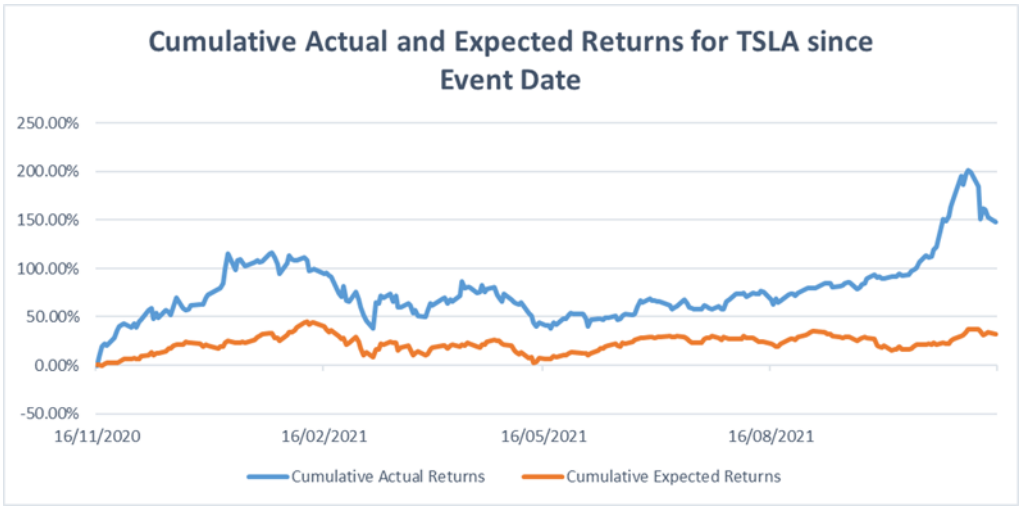
- delnica TSLA je presegla pričakovanja za 226,50 %,
- delnica GME je presegla pričakovanja za 2030,59 %,
- delnica AMC je presegla pričakovanja za 6345,30 %,
- delnica PLTR je presegla pričakovanja za 47,56 %,
- delnica NIO je presegla pričakovanja za 173,38 %.

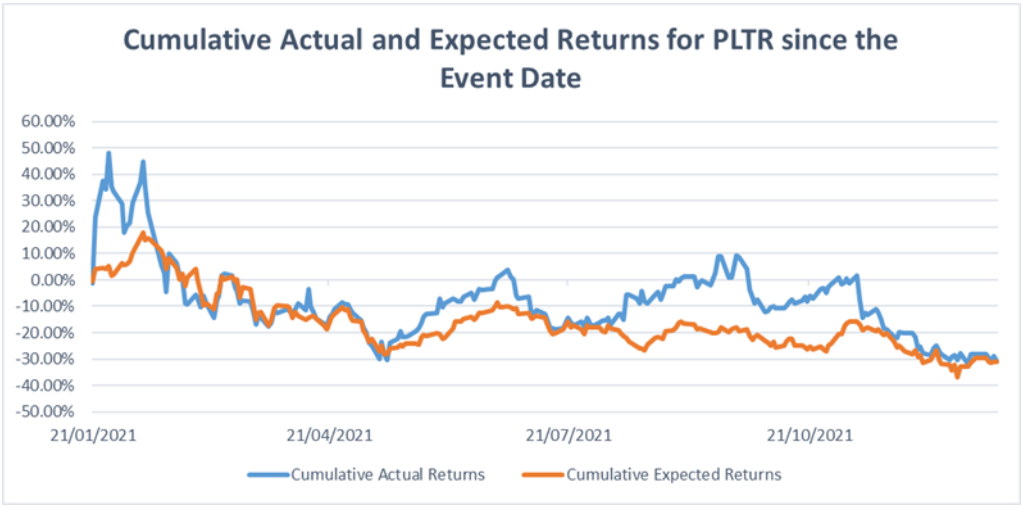
Potem ko smo ugotovili, da je vseh pet delnic preseglo pričakovanja, smo izvedli analizo, da bi ugotovili, ali je ta presežek mogoče pripisati razpravam na Redditu in Twitterju. To smo storili tako, da smo izračunali korelacijo in opravili regresije med dnevnimi omembami simbolov delnic na družbenih omrežjih in dnevnimi absolutnimi donosi, ki so jih uspele doseči, pri čemer so bili rezultati naslednji:

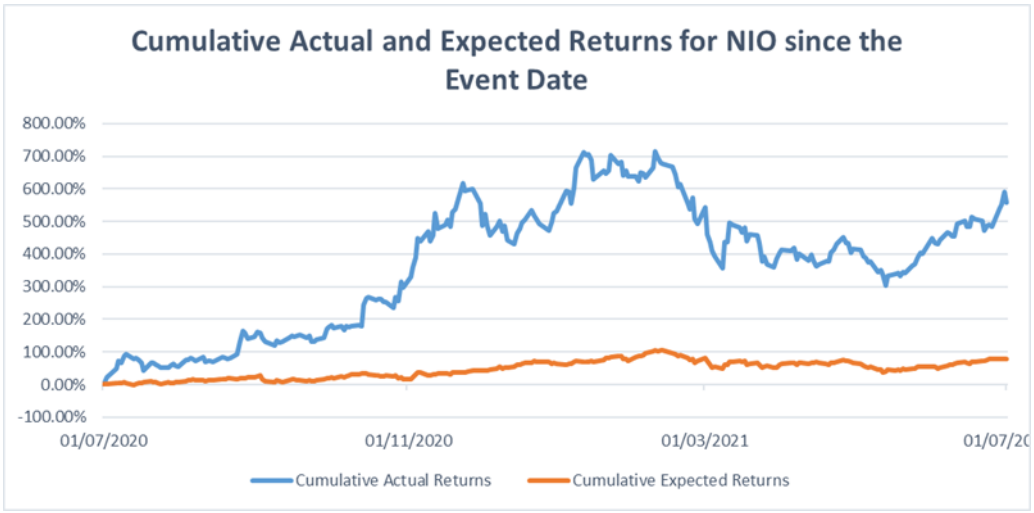
- Delnica TSLA je pokazala korelacijo 0,42 in 0,40 z dnevnimi omembami na Redditu oziroma Twitterju; 17,8 % njenih absolutnih donosov je pojasnjenih z razpravami na Redditu, 16,1 % pa z razpravami na Twitterju.
- Delnica GME je pokazala korelacijo 0,73 in –0,06 z dnevnimi omembami na Redditu oziroma Twitterju; 54,0 % njenih absolutnih donosov je pojasnjenih z razpravami na Redditu, 0,31 % pa z razpravami na Twitterju.
- Delnica AMC je pokazala korelacijo 0,72 in 0,24 z dnevnimi omembami na Redditu oziroma Twitterju; 51,6 % njenih absolutnih donosov je pojasnjenih z razpravami na Redditu, 5,5 % pa z razpravami na Twitterju.
- Delnica PLTR je pokazala korelacijo 0,57 in 0,23 z dnevnimi omembami na Redditu oziroma Twitterju; 32,0 % njenih absolutnih donosov je pojasnjenih z razpravami na Redditu, 5,1 % pa z razpravami na Twitterju.
- Delnica NIO je pokazala korelacijo 0,15 in 0,45 z dnevnimi omembami na Redditu oziroma Twitterju; 2,1 % njenih absolutnih donosov je pojasnjenih z razpravami na Redditu, 20,6 % pa z razpravami na Twitterju.

Na podlagi zgornjih rezultatov lahko zaključimo, da je za delnice TSLA, GME, AMC in PLTR med 17 in 55 % njihovih absolutnih donosov mogoče pojasniti z razpravami na Redditu, korelacija med njihovo ceno delnic in obsegom razprav pa je pomembna. TSLA in NIO kažeta pomembno korelacijo z omembami na Twitterju, približno 20 % njunih absolutnih donosov pa je mogoče pojasniti z razpravami na platformi.

Appendix 2: Cumulative actual and expected returns for TSLA, PLTR, and NIO since the event date







Appendix 3: Alternative Fama French 3 factor models for TSLA, PLTR, and NIO

TSLA

Regression Statistics	
Multiple R	0.57155
R Square	0.32667
Adjusted R Square	0.31853
Standard Error	0.02931
Observations	252

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	3	0.10338	0.03446	40.10644	0.00000
Residual	248	0.21308	0.00086		
Total	251	0.31645			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	0.00291	0.00187	1.55192	0.12196	-0.00078	0.00659
Mkt-RF	0.01369	0.00248	5.52241	0.00000	0.00880	0.01857
SMB	0.01031	0.00240	4.30350	0.00002	0.00559	0.01502
HML	-0.00973	0.00168	-5.79700	0.00000	-0.01303	-0.00642

PLTR

Regression Statistics	
Multiple R	0.58344
R Square	0.34040
Adjusted R Square	0.33202
Standard Error	0.03386
Observations	240

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	3	0.13960	0.04653	40.59825	0.00000
Residual	236	0.27050	0.00115		
Total	239	0.41010			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	0.00056	0.00221	0.25496	0.79898	-0.00379	0.00492
Mkt-RF	0.00308	0.00275	1.12255	0.26277	-0.00233	0.00850
SMB	0.02428	0.00288	8.43482	0.00000	0.01861	0.02995
HML	-0.00914	0.00200	-4.56611	0.00001	-0.01309	-0.00520

(table continues)

(continued)

NIO

Regression Statistics	
Multiple R	0.46271
R Square	0.21410
Adjusted R Square	0.20464
Standard Error	0.05118
Observations	253

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	3	0.17767	0.05922	22.61194	0.00000
Residual	249	0.65216	0.00262		
Total	252	0.82983			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	0.00640	0.00326	1.95948	0.05117	-0.00003	0.01283
Mkt-RF	0.01534	0.00349	4.39696	0.00002	0.00847	0.02220
SMB	0.01280	0.00399	3.20732	0.00152	0.00494	0.02066
HML	-0.01073	0.00242	-4.43038	0.00001	-0.01550	-0.00596