2022

The Rising Power of the Individual Investor: How Social Media Sentiments and User Activity Impact Stock Price Volatility and Trading Volume

Nandini Jayaram
Claremont McKenna College

Claremont McKenna College

The Rising Power of the Individual Investor: How Social Media Sentiments and User Activity Impact Stock Price Volatility and Trading Volume

Submitted to
Professor Dass

By
Nandini Jayaram

For
Senior Thesis
Fall 2021
December 6, 2021

Acknowledgements

I would like to sincerely thank my advisor, Professor Nishant Dass, and mentor, Professor Yong Kim, for their continued support and patience. Through Professor Kim's "Research Methods" course, I was given a structured timeline to follow throughout the semester, and would not have been able to make consistent progress without the feedback and participation from the class. I would also like to express gratitude to Claremont McKenna College for giving me the opportunity and resources to explore a topic of my choice, and culminate my academic career with this thesis.

A special thanks to Professor Marc Massoud and Professor Peter McAniff for their guidance and invaluable mentorship over the years, which strengthened my deep interest in finance. Professor McAniff's "Financial Statement Analysis" class, where we discussed global financial news on a daily basis, such as the AMC and GME short squeezes, was particularly inspirational and caused me to further explore this topic.

I am also extremely grateful to my parents for supporting me and my education, and my friends, especially Shreya, Vishwa, Nick, and Dini for offering encouragement and motivating me throughout this process.

Abstract

This paper investigates the relationship between social media sentiments and user activity on the ability of online platforms like Robinhood, Reddit, and Twitter to drive volatility in stock returns and trading volume of shares. The research contributes to the existing literature by studying the potential power of social media communities to influence the stock market parameters. The dataset includes information for several stocks of daily Robinhood users, mentions in r/wallstreetbets, number of Twitter followers, as well as closing price and trading volume data. Using a fixed effects panel data model, the regressions yield statistically significant results that indicate higher WallStreetBets mentions and a larger number of Robinhood users are positively associated with increases in price and total volume, while more Twitter followers are positively associated with lower closing prices. Additionally, regression results show that more Robinhood users are associated with lower stock price volatility, while increased Twitter followers are associated with higher stock price volatility. This paper uniquely studies the impact of investment-related sentiments and user activity across three social media platforms, and offers a new finding that identifies the positive association between Twitter followers and stock price volatility. Future studies can examine a larger portfolio of stocks over a longer time period to establish a stronger relationship between social media signals on stock return volatility and trading volume.

Table of Contents

1. Introduction

In the last decade, Wall Street has seen a shifting balance of power from large traditional hedge funds and professional managers to individual retail investors. Through stock-trading mobile apps such as Robinhood, anyone can now purchase fractional shares of a stock, and they don't have to pay transaction fees to buy or sell shares (Heath 2020). Social media platforms like Reddit enable users to discuss their investing opinions in online communities, such as a subreddit group like r/wallstreetbets. As a result, the playing field has been leveled whereby the common investor has increased power to influence the markets. Both of Robinhood and Reddit platforms were created between early 2012-2013 and saw rapid growth during the COVID-19 pandemic, partly due to support from political and celebrity figures (Ingram and Abbruzzese 2021). Today, the r/wallstreetbets subreddit has almost 11 million subscribers, and Robinhood with its community of almost 23 million users has been rewarded with a valuation of $32 billion after its IPO in July 2021 (Fitzgerald and Levy 2021). Especially in 2021, since the GameStop short squeeze has occurred, more attention and research has been conducted on the impact of these social media factors on the stock market.

This paper examines the predicted association between both social media sentiments and Robinhood user activity on stock market data through a comparative analysis. Based on existing literature, I hypothesize that both variables (increased social media sentiments and higher Robinhood user activity) will have an impact on stock prices and trading volume. I believe that this hypothesis is worth exploring because the results can shed light on whether social media interactions among millennial investors have the ability to increase stock return volatility and trading volume of stocks that are popular with retail investors. If a pattern can be established that investors will make decisions based on social media signals and rely on

Reddit, Robinhood, and Twitter communities to inform their trades, then there is room for further research on an increase in noise trading leading to more informed trading. **To meaningfully contribute to this area of research, this paper uniquely studies the impact of sentiments and user activity across three social media platforms, and offers a new finding that identifies the positive association between Twitter followers and stock price volatility.** The results also find that higher WallStreetBets mentions and a higher number of Robinhood users are positively associated with increases in closing price and total volume, while more Twitter followers are positively associated with lower closing prices. **These empirical findings support my prediction that the effects of social media sentiments will enable the individual investor to drive volatility in stock prices and total volume of shares, and thus change the future of trading.**

The rest of the paper is organized into the following parts: Section 2 provides an overview of existing literature; Section 3 covers data collection and research methods; Section 4 includes the empirical analysis of regression results and limitations of the study; Section 5 summarizes the findings and offers a conclusion; Section 6 lists the references used, and Section 7 includes tables in the appendix.

2. Literature Review

There is existing literature that uses data from Reddit or Robinhood user activity to analyze the impact of attention-induced trading on stock returns. "Bet it on Reddit: The Effects of Reddit Chatter on Highly Shorted Stocks" uses WallStreetBets as a quantifiable case study to examine social media users' potential causal impact on highly shorted stocks. Rather than focusing on the predictive power of sentiment analysis, it looks at the potential power of social media communities to affect the market. The study yields statistically significant findings that focused attention from a social media platform can shift and organize stock prices (Diangson and Jung 2021). Barber, Huang, Odean, and Schwarz (2021) find in "Attention-Induced Trading and Returns: Evidence from Robinhood Users" that increased speculative trading by Robinhood investors, who also commonly use channels like WallStreetBets to share market information, further contributes to attention-driven trading. By studying "herding episodes," which the authors define as days that record a significant increase of a group of Robinhood users holding a specific stock, they predict that waves of intense buying will forecast large negative returns. The paper finds that attention-driven trading on a small portfolio of stocks can develop a herding mentality and result in negative abnormal returns.

In contrast to these findings, another paper, "Zero-Commission Individual Investors, High Frequency Traders, and Stock Market Quality," by Eaton, Green, Roseman, and Wu (2021), finds no association between changes in Robinhood ownership and future predicted returns. The authors conclude that zero-commission investors, such as individuals who use Robinhood's commission-free trading model, act as noise traders and negatively affect the quality of the financial market. By analyzing mentions on WallStreetBets, the authors find that stocks preferred by Robinhood traders are correlated with higher market liquidity and decreased return volatility. Similarly, in "Individual Investor Trading and Stock Returns," Kaniel, Saar, and Titman (2008) study the dynamic interaction between short-horizon returns and individual investors by looking at periods of intense buying or selling. While the nature of the data does not enable the authors to make conclusions about the long-term impact of individual trading on sales, the paper's findings support patterns from the literature that intense buying is followed by positive excess returns and intense selling is followed by negative excess returns.

A related area of literature that is relevant to this research topic is the effect of social media platforms on price dynamics. "The Rise of Reddit: How Social Media Affects Retail Investors and Short-sellers' Roles in Price Discovery" by Hu, Jones, Zhang, and Zhang (2021), investigates the effect of four components of social media activity on stock prices, retail order flows, and short-selling. The results support that the four measures of Reddit activity, including higher Reddit traffic, positive tone, more disagreement, and increased network connectedness are correlated with higher stock returns, positive retail order flow, and decreased shorting flows in the future. The authors also find that increased Reddit traffic strengthens the positive predictive power of retail order flow, and that all these results predict even lower stock returns. When there is more social media activity, retail buying behavior is encouraged and shorting is discouraged, as short sellers tend to be more risk-averse to avoid short squeezes, and they use social media as an informative tool to predict future negative returns.

Another area of literature that focuses on social media sentiments includes the paper, "'I Just Like the Stock' versus 'Fear and Loathing on Main Street': The Role of Reddit

Sentiment in the GameStop Short Squeeze" by Long, Lucey, and Yarovaya (2021). This study measures the role of investment related sentiments from Reddit, through tone and timing of posts on WallStreetBets, on the GameStop stock price changes. To specify the feelings conveyed in investor sentiments from Wall Street Bets, the authors categorize emotions from posts into "Angry," "Fear," "Happy," "Sad," and "Surprise." They find a strong association between number of comments posted in WallStreetBets per hour and GameStop prices. From the sentiment analysis, results show that "Fear" is the most common sentiment and "Sad" has the most significant effect on GameStop 1-minute returns, while sentiments from longer threads have more impact on intraday returns than shorter threads.

My research paper will add to the existing literature that studies the impact of investment-related social media sentiments on the stock market. By conducting an empirical analysis that examines the association of social media sentiments and Robinhood user data on stock price volatility and trading volume of shares, I seek to contribute to literature that has already studied the separate effects of these variables on the market. The research in this area is relatively new and therefore, findings from existing papers are inconclusive, especially regarding the impact of intense buying by Robinhood investors (some papers find that intense buying predicts positive returns, some predict negative returns, and some find no association at all). By analyzing the effect of these variables separately on a selected group of stocks, the results from my empirical analysis show that Robinhood users have a large effect on stock price volatility, and mentions on WallStreetBets and Twitter followers are positively correlated with closing price and volume.

3. Data and Research Methods

Sample Selection

The purpose of this empirical analysis is to investigate if there is a relationship between a higher number of millennial investors participating in trades and increased volatility in stock prices and volume, with a specific focus on Robinhood user activity, Reddit, and Twitter sentiment analysis. My research focuses on identifying an association between stock price and volume movements, and the behaviors of noise traders over time. The dependent variables in my analyses are closing price and total volume of shares, and I investigate the relationships between price/volume and four independent variables (Robinhood users, WallStreetBets (WSB) mentions, Twitter follower count, and total volume of shares).

I created a dataset comprising a portfolio of seven stocks: Apple Inc. (AAPL), AMC Entertainment Holdings Inc. (AMC), Amazon.com Inc. (AMZN), Meta Platforms Inc. (FB), GameStop Corp. (GME), Microsoft Corporation (MSFT), and Tesla Inc. (TSLA). I chose this combination of stocks to determine if noise trading behaviors could drive stock price and volume volatility in Big Tech stocks as well as the popular highly shorted stocks that received heavy media coverage such as GameStop and AMC. For each stock ticker, I collected daily data on Robinhood users holding each stock, market data on closing prices and total volume, and the number of WSB mentions and Twitter followers in the period between August 1, 2018, and October 29, 2021. The information on Robinhood user activity comes from Robintrack, a website that posts updates of stock popularity on Robinhood, but as of August 2020, Robinhood discontinued the API that allowed the site to collect popularity data. This data was listed on an hourly basis, every day from May 2, 2018, to August 13, 2020. In order to capture the last updated count of Robinhood users per day, I manually cleaned the data in order to find the final value recorded each day in that time period for all seven stocks. The market data on each stock was pulled from Yahoo Finance's repository of historical financial data. Finally, the number of mentions in WSB and followers on Twitter is sourced from Quiver Quantitative, a data platform that compiles market and company information to aid with investment analysis. This website includes many datasets that are updated daily, ranging from Senate and House trades to insider trading to the most-mentioned stocks in the WSB forum.

Due to the varying timelines and dates available for the different variables (i.e., Robinhood user data is unavailable after August 2020, while market data, WSB mentions, and Twitter followers is up to date) I created one cleaned dataset that standardizes the overlapping time period across all variables for each of the seven stocks and included the extended timeline until October 2021. While I was unable to include data on Robinhood users after August 2020, I believe that my analysis of daily data over this three-year timeframe will help establish a predicted association that can be applied in future research. Descriptive summary statistics of the data used for my analysis can be found in Tables A1-4 in the Appendix. Table A1 shows the average, standard deviation, minimum and maximum values for all variables. For example, WSB mentions range from 1 to 19,446 with an average of 185 mentions. Tables A2, A3, and A4 show the average, standard deviation, and frequency of each of the three variables (closing prices, WSB mentions, and Robinhood users) sorted by each of the seven stocks to analyze the trends of these variables over time. In Table A3, we can see that the average values of WSB mentions are highest for the two heavily shorted stocks: AMC has a mean of 320 mentions and GME has a mean of 564 mentions. Table A4 shows that GME has the lowest average of Robinhood users at 16,103 compared to Big Tech stocks, like AAPL with an average of 259,350 users and MSFT with an average of 257,700 users.

Design

The summary statistics and trends over time represent the need for further empirical analysis to understand the exact relationships between stock price and these variables. My model includes time series panel data to measure the changes in Robinhood users, number of WSB mentions, Twitter followers, and total volume of shares over a three-year timeframe for each stock. As the function for panel data cannot include string variables such as ticker symbols, I generated a new variable, *StockName*, as the numeric substitute for the string variable, *StockTicker*. Once that was adjusted, I established my panel variable as *StockName*, which categorizes the data collected into the seven stocks that are being analyzed, and the time variable as *Date* to specify the order of observations within the panel, with a delta unit of 1 day, the periodicity of the time variable. I decided to use a fixed effects model to remove any possible bias that could correlate independent variables in the dataset with time-invariant, unmeasured variables.

As further modifications to certain variables, I created *PriceChange* as a measure of daily change in closing price rather than simply observing the effects of these variables on *ClosingPrice* and used the same equations to create *RHChange* and *TwitterChange* to measure a similar day-over-day change in Robinhood users and Twitter followers. I also created logarithmic transformations for total volume and Twitter followers to adjust for especially high numeric values which could skew the results. Following these adjustments, I ran four panel data regressions with *ClosingPrice*, *PriceChange*, *TotalVolume*, and *logVolume* as the dependent variables. I first conducted correlational tests between *ClosingPrice* and the four independent variables (*WSBMentions, TotalVolume, TwitterFollowers, and RHUsers)*. Due to the lack of time overlap between Robinhood users and Twitter followers for GME and AMC, I ran additional tests for those two stocks to identify correlations between *ClosingPrice*, *RHUsers*, and *TotalVolume*. As seen below in Tables 1 and 2, *TotalVolume* has a strong positive correlation with *WSBMentions* (0.89 for AMC and 0.93 for GME), while *RHUsers* has a strong negative correlation with *ClosingPrice* (-0.63 for AMC and -0.72 for GME). Figure 1 depicts the negative relationship between closing price and Robinhood users for GameStop. These correlation results potentially indicate that negative sentiments on Robinhood impact price while WSB mentions drive increased trading volume for AMC and GME.

Table 1: AMC matrix of correlations

Variables	(1)	(2)	(3)	(4)
(1) ClosingPrice	1.000			
(2) WSBMentions	0.119	1.000		
(3) TotalVolume	0.384	0.887	1.000	
(4) TwitterFollowers	0.148	-0.434	-0.333	1.000

Variables	(1)	(2)	(3)
(1) ClosingPrice	1.000		
(2) RHUsers	-0.635	1.000	
(3) TotalVolume	-0.335	0.353	1.000

Table 2: GME matrix of correlations

Variables	(1)	(2)	(3)	(4)
(1) ClosingPrice	1.000			
(2) WSBMentions	0.443	1.000		
(3) TotalVolume	0.509	0.926	1.000	
(4) TwitterFollowers	0.282	-0.123	-0.084	1.000

Variables	(1)	(2)	(3)
(1) ClosingPrice	1.000		
(2) RHUsers	-0.721	1.000	
(3) TotalVolume	-0.187	-0.012	1.000

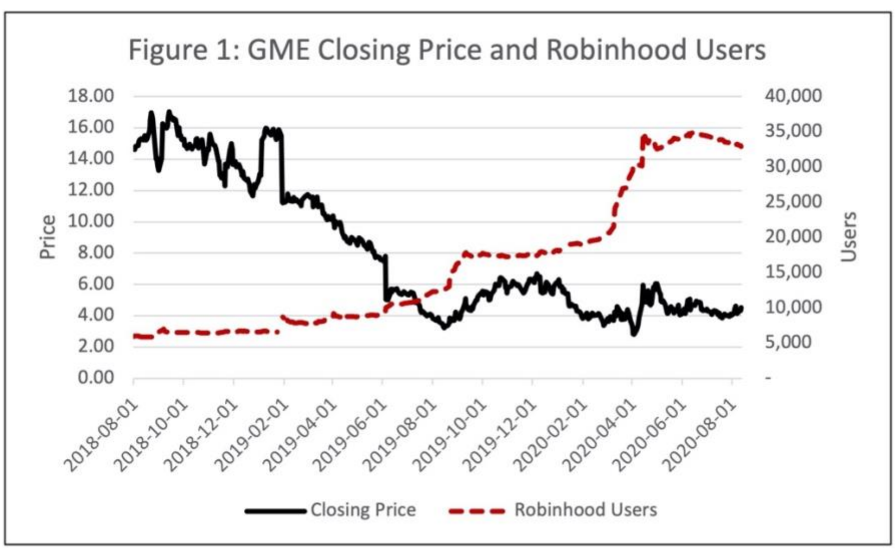

Figure 1: GME Closing Price and Robinhood Users

On the other hand, Tables 3-7 show that for AAPL, AMZN, MSFT, and TSLA, *RHUsers* and *TwitterFollowers* have strong positive correlations with *ClosingPrice*, and there are strong positive correlations in general between *TwitterFollowers* and *RHUsers*. An exception to this pattern is FB, which shows a negative correlation between *TwitterFollowers* and *ClosingPrice* (-0.76), and a strong negative correlation between *RHUsers* and *TwitterFollowers* (-0.92). There is a moderately strong correlation of 0.62 between *WSBMentions* and *ClosingPrice* for AAPL, and as seen below in Figure 2, there is a large spike between August and October 2020, reflecting the maximum number of AAPL mentions (1,970) in September 2020. However, we see that Twitter followers and Robinhood users have weak correlations with total volume of shares. These results show different patterns between the two highly shorted stocks versus Big Tech stocks, specifically the relationships between Robinhood users and closing price, but similar trends with the weak correlations between Twitter followers, Robinhood users, and total volume.

Figure 2: AAPL Closing Price and WSB Mentions

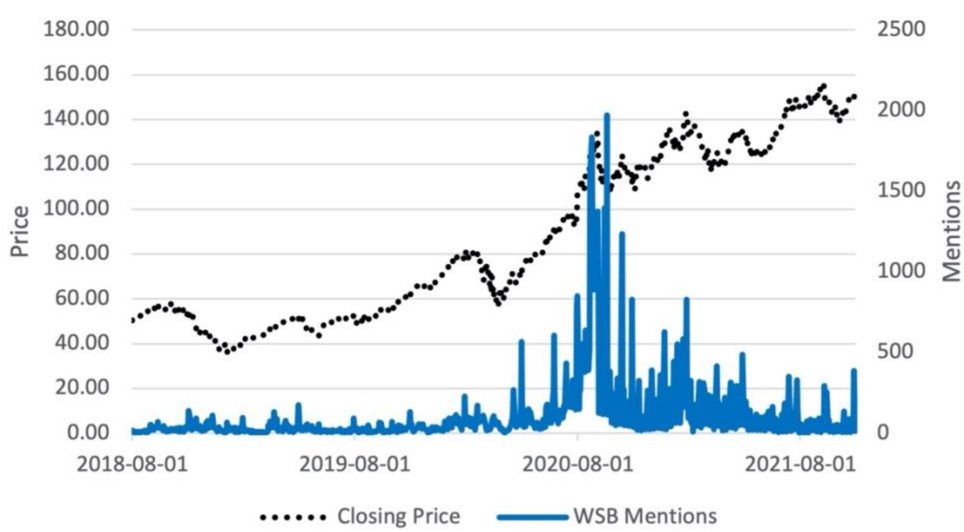

Table 3: AAPL matrix of correlations

Variables	(1)	(2)	(3)	(4)	(5)
(1) ClosingPrice	1.000				
(2) WSBMentions	0.620	1.000			
(3) RHUsers	0.970	0.655	1.000		
(4) TotalVolume	0.132	0.419	0.153	1.000	
(5) TwitterFollowers	0.970	0.558	0.964	0.040	1.000

Table 4: AMZN matrix of correlations

Variables	(1)	(2)	(3)	(4)	(5)
(1) ClosingPrice	1.000				
(2) WSBMentions	0.431	1.000			
(3) RHUsers	0.952	0.316	1.000		
(4) TotalVolume	0.001	0.716	-0.099	1.000	
(5) TwitterFollowers	0.900	0.212	0.845	-0.195	1.000

Table 5: FB matrix of correlations

Variables	(1)	(2)	(3)	(4)	(5)
(1) ClosingPrice	1.000				
(2) WSBMentions	0.178	1.000			
(3) RHUsers	0.753	0.104	1.000		
(4) TotalVolume	0.081	0.733	0.139	1.000	
(5) TwitterFollowers	-0.761	0.048	-0.923	0.028	1.000

Table 6: MSFT matrix of correlations

Variables	(1)	(2)	(3)	(4)	(5)
(1) ClosingPrice	1.000				
(2) WSBMentions	0.405	1.000			
(3) RHUsers	0.868	0.384	1.000		
(4) TotalVolume	-0.042	0.374	0.088	1.000	
(5) TwitterFollowers	0.865	0.325	0.944	0.033	1.000

Table 7: TSLA matrix of correlations

Variables	(1)	(2)	(3)	(4)	(5)
(1) ClosingPrice	1.000				
(2) WSBMentions	0.491	1.000			
(3) RHUsers	0.966	0.388	1.000		
(4) TotalVolume	0.026	0.681	-0.028	1.000	
(5) TwitterFollowers	0.936	0.338	0.927	-0.140	1.000

4. Empirical Analysis

Fixed Effects Panel Data Model

This paper implements a fixed effects panel data model with four variations to assess the predicted association of social media variables on closing price and volume. The first regression uses the following equation:

$$\boldsymbol{Closing\ Price}_{(\mathrm{i,t})} = \boldsymbol{\alpha}_{\mathrm{i}} + \boldsymbol{\beta}_1 \boldsymbol{WSB}_{\mathrm{it}} + \boldsymbol{\beta}_2 \boldsymbol{RH}_{\mathrm{it}} + \boldsymbol{\beta}_3 \boldsymbol{Twtr}_{\mathrm{it}} + \boldsymbol{\beta}_4 \boldsymbol{Volume}_{\mathrm{it}} + \boldsymbol{\varepsilon}_{\mathrm{it}} \quad \textbf{(1)}$$

In Equation 1, *ClosingPrice* is the dependent variable for daily stock market data with *i* denoting each stock ticker and *t* denoting year. α_{i} represents the level of significance and fixed effects, $\beta_1 WSB_{\mathrm{it}}$ represents the impact of WallStreetBets mentions, $\beta_2 RH_{\mathrm{it}}$ represents the impact of Robinhood users, $\beta_3 Twtr_{\mathrm{it}}$ represents the impact of Twitter followers, $\beta_4 Volume_{\mathrm{it}}$ represents the total volume of shares, and $\varepsilon_{\mathrm{it}}$ represents the error term. This regression maintains the original values of all variables, which could result in high values for Twitter followers and total volume skewing the normalized dataset. Table 8 yields statistically significant p-values for Robinhood users, Twitter followers, and total volume at the 1% significance level, and WSB mentions at the 10% significance level. Holding all else fixed, with one additional unit increase in *WSBMentions* and one additional unit increase in *RHUsers*, there is an associated increase in *ClosingPrice* of 0.05 and 0.002, respectively. With an additional unit increase in *TwitterFollowers*, there is an associated decrease in *ClosingPrice* of 0.001.

Table 8: Regression results for Closing Price

ClosingPrice	Coef.	St.Err.	t-value	p-value	[95% Conf	Interval]
WSBMentions	0.05 *	0.028	1.78	0.076	-0.005	0.104
RHUsers	0.002 ***	0	15.71	0	0.001	0.002
TwitterFollowers	-0.001 ***	0	-8.42	0	-0.001	0
TotalVolume	0 ***	0	-2.62	0.009	0	0
Constant	4,098.3***	446.588	9.18	0	3219.839	4,976.761
R-squared			0.480	Number of obs		345

**** p<.01, ** p<.05, * p<.1*

The second regression, which uses a modified dependent variable, *PriceChange*, to reflect the difference in closing price between two consecutive days, is depicted by the following equation:

$$\boldsymbol{Price\ Change}_{(i,t)} = \boldsymbol{\alpha_i} + \boldsymbol{\beta_1 WSB_{it}} + \boldsymbol{\beta_2 RHChange_{it}} + \boldsymbol{\beta_3 TwtrChange_{it}} + \boldsymbol{\beta_4 Volume_{it}} + \boldsymbol{\varepsilon_{it}} \quad \textbf{(2)}$$

In Equation 2, α_i represents the level of significance and fixed effects with *i* denoting each stock ticker and *t* denoting year, $\beta_1 WSB_{it}$ represents the impact of WallStreetBets mentions, $\beta_2 RHChange_{it}$ represents the daily change in Robinhood users, $\beta_3 TwtrChange_{it}$ represents the daily change in Twitter followers, $\beta_4 Volume_{it}$ represents the total volume of shares, and ε_{it} represents the error term. This regression includes modifications of the dependent variable, *PriceChange* and two independent variables, *RHChange* and *TwitterChange*. Table 9 shows statistically significant p-values for number of WSB mentions and change in Twitter followers at the 1% significance level and change in Robinhood users at the 5% significance level. Holding all else fixed, with one additional unit change in *RHUsers*, there is an associated decrease of change in *ClosingPrice* of 0.422 and with one additional unit change of *TwitterFollowers*, there is an associated increase of change in *ClosingPrice* of 3.958.

Table 9: Regression results for Price Change

PriceChange	Coef.	St.Err.	t-value	p-value	[95% Conf	Interval]
WSBMentions	0***	0	2.97	0.003	0	0
RHChange	-0.422**	0.191	-2.21	0.028	-0.799	-0.046
TwitterChange	3.958***	1.4	2.83	0.005	1.2	6.716
TotalVolume	0	0	-0.48	0.634	0	0
Constant	0.001	0.004	0.16	0.872	-0.008	0.01
R-squared			0.059	Number of obs		257

**** p<.01, ** p<.05, * p<.1*

The third regression switches the dependent variable to total volume of shares, and measures the effect of closing price, WSB mentions, Robinhood users, and Twitter followers on volume. This regression uses the following equation:

$$\boldsymbol{Total\,Volume_{(i,t)} = \alpha_i + \beta_1 ClPrice_{it} + \beta_2 WSB_{it} + \beta_3 RH_{it} + \beta_4 Twtr_{it} + \varepsilon_{it}} \tag{3}$$

In Equation 3, α_i is the level of significance and fixed effects, *i* denotes each stock ticker and *t* denotes year, $\beta_1 ClPrice_{it}$ is the impact of closing price, $\beta_2 WSB_{it}$ is the impact of WSB mentions, $\beta_3 RH_{it}$ is the impact of Robinhood users, $\beta_4 Twtr_{it}$ is the impact of Twitter followers, and ε_{it} is the error term. This regression includes the original format of total volume, which includes high numeric values and result in moderately skewed results. Table 10 shows statistically significant p-values for closing price, WSB mentions, Robinhood users, and Twitter followers at the 1% significance level. The coefficients for closing price and WSB mentions are very high numbers because of the skew resulting from using original total volume data. Holding all else fixed, with one additional unit increase in *RHUsers*, there is an associated increase in *TotalVolume* of 71.594 and with one additional unit increase in *TwitterFollowers*, there is an associated decrease in *TotalVolume* of 60.924. Every additional *WSBMention* is associated with an increase in *TotalVolume* of 45,581.138 and with one unit increase in *ClosingPrice*, there is an associated decrease in *TotalVolume* of 26,008.28, suggesting its outsized impact.

Table 10: Regression results of Total Volume

TotalVolume	Coef.	St.Err.	t-value	p-value	[95% Conf	Interval]
ClosingPrice	-26,008.28***	9,915.235	-2.62	0.009	-45,512.037	-6,504.52
WSBMentions	45,581.138***	4,495.754	10.14	0	36,737.768	54,424.508
RHUsers	71.594***	25.26	2.83	0.005	21.905	121.282
TwitterFollowers	-60.924***	12.836	-4.75	0	-86.173	-35.675
Constant	4.741e+08***	87,971,706	5.39	0	3.011e+08	6.472e+08
R-squared			0.265	Number of obs		345

**** p<.01, ** p<.05, * p<.1*

Finally, the fourth regression modifies total volume and uses *logVolume* as the dependent variable, to adjust for the skewness toward high values that comes from the original total volume and removes the concern of outliers. This regression is estimated by the following equation:

$$\boldsymbol{log\ Volume_{(i,t)} = \alpha_i + \beta_1 PriceChange_{it} + \beta_2 WSB_{it} + \beta_3 RH_{it} + \beta_4 Twtr_{it} + \varepsilon_{it}} \qquad (4)$$

In Equation 4, α_i represents the level of significance and fixed effects, *i* denotes each stock ticker and t denotes year, $\beta_1 PriceChange_{it}$ represents the impact of the change in closing price, $\beta_2 WSB_{it}$ represents the impact of WSB mentions, $\beta_3 RH_{it}$ represents the impact of Robinhood users, $\beta_4 Twtr_{it}$ represents the impact of Twitter followers, and ε_{it} represents the error term. This regression includes the original format of all independent variables except for price, and logarithmically transforms the total volume to use *logVolume* as the dependent variable. Table 11 shows statistically significant p-values for WSB mentions and Twitter followers at the 1% significance level. Holding all else fixed, with one additional WSB mention, there is an associated increase in volume of shares by 0.1%.

Table 11: Regression results of logVolume

logVolume	Coef.	St.Err.	t-value	p-value	[95% Conf	Interval]
PriceChange	0.444	0.508	0.87	0.383	-0.555	1.443
WSBMentions	0.001***	0	13.93	0	0.001	0.001
RHUsers	0	0	0.38	0.707	0	0
TwitterFollowers	0 ***	0	-4.82	0	0	0
Constant	22.054***	0.981	22.49	0	20.125	23.983
R-squared			0.400	Number of obs		345

**** p<.01, ** p<.05, * p<.1*

Across these four regression results, the first key takeaway is that *RHUsers* are insignificant when *logVolume* is the dependent variable, while *RHChange* and *RHUsers* are statistically significant when total volume and a variation of closing price are the dependent variables. This indicates that Robinhood users have a larger effect on stock prices rather than the volume of shares, but according to Table 10, with one additional unit increase in Robinhood users, there is an associated increase in total volume of 71.594. Next, *WSBMentions* and *TwitterFollowers* are significant when both price and volume are dependent variables. This indicates that WSB mentions and Twitter followers have a strong positive correlation with increasing both stock price and volume of shares. Finally, the relationship between closing price and volume of shares is significant when the original variables are maintained, but becomes insignificant when the dependent variable is modified (*PriceChange* and *logVolume*).

Each of the regressions yields relatively low coefficients of determination, or R-squared values. The first regression (using *ClosingPrice* as the dependent variable) has the highest R-squared value of 0.480. The second regression (using *PriceChange* as the dependent variable) shows the lowest R-squared value of 0.059. The third regression (using *TotalVolume* as the dependent variable) has a R-squared value of 0.265. The fourth regression (using *logVolume* as the dependent variable) has a R-squared value of 0.400. The R-squared value from the first regression indicates that only 48% of the variance of *ClosingPrice* can be depicted by the variance of the four independent variables. However, since this empirical analysis studies social media chatter and user activity driven by human behavior, the low R-squared values should not outweigh the p-values which indicate the level of marginal significance in this test. We can identify positive predicted values of WSB mentions on closing price and volume response values, and positive and negative predicted values of Robinhood users and Twitter followers on price and volume response values. This can be explained by the missing data for Robinhood users and Twitter followers, while WSB mentions, price, and volume data is complete for almost the entire time period.

Additional Considerations and Limitations

The purpose of this section is to account for limitations of this fixed effects panel data analysis and recommend considerations for future research. In this study, only a small selection of seven stocks were analyzed, including two highly volatile and shorted stocks that gained media coverage in early 2021, and five Big Tech stocks that consistently garner media attention. Future papers can expand on this portfolio of stocks that are analyzed. Additionally, because the data for Robinhood users is unavailable after August 2020, future studies can design scraping programs to find more updated and accessible data on Robinhood user activity, potentially even information on types and volume of stock trades made by Robinhood users on a daily basis.

This study examined the variables over a three-year timeframe, from August 2018 to October 2021, which includes the volatile state of the stock market that has resulted from the COVID-19 pandemic. Since the effects of subreddit communities, Twitter followers, and retail investors using Robinhood are very recent and therefore difficult to establish conclusions, future research could implement a larger timeframe to measure any significant effects of social media activity on stock prices and volume. This paper does not account for temporary spikes of closing price and share volume in response to quarterly earnings announcements or increased attention from news coverage, which I predict would create substantial reactions from noise traders responding to a surprise in company earnings and media.

5. Conclusion

This paper attempts to establish a predicted association between Robinhood users, mentions on the WallStreetBets forum, and number of Twitter followers on stock price volatility and trading volume of shares. The data for this research was collected from Robintrack, Yahoo Finance, and Quiver Quantitative, and I used these resources to create a dataset for seven stocks including daily values for each variable across a three-year timeframe from August 1, 2018, to October 29, 2021. After finding summary statistics and creating visualizations which emphasized important correlations between Robinhood users, WSB mentions, and closing price, I conducted four panel data regressions using variations of closing price and total volume as the dependent variables.

My analyses found that social media forums impact price volatility and trading volume differently. First, the results from the first and third regressions show that higher WSB mentions and a larger number of Robinhood users are positively associated with increases in closing price and total share volume. On the other hand, more Twitter followers are associated with lower closing prices. While these findings seem to align well with the conclusion from Diangson and Jung (2021) that focused attention from social media platforms can shift stock prices, my results are nuanced by the negative association between Twitter followers and closing prices, possibly indicating that changes in Twitter users following or unfollowing these companies negatively affects closing prices.

Similarly, the results from the second regression also demonstrate opposite impacts on price volatility between Robinhood and Twitter forums. Specifically, a higher number of Robinhood users is associated with lower stock price volatility, while a higher number of

Twitter followers is associated with higher stock price volatility. While the findings regarding Robinhood user activity is supported by Eaton et al. (2021) which finds that stocks preferred by Robinhood traders are correlated with decreased return volatility**, my research offers a new finding that associates increased Twitter followers with higher price volatility. Overall, these results suggest that my initial hypothesis (increased social media sentiments and higher Robinhood user activity will impact stock prices and trading volume) is correct and provides evidence that social media enables retail investors to drive volatility in stock prices and trading volume.**

While this research only establishes a pattern for seven stocks in a short timeframe, there is room for future studies to measure the predicted impact of noise trading because of signals from social media communities. The results from this study can inspire additional research that considers many other factors which could affect price and volume of stocks. These factors can include if social media sentiments are stronger in locations that have closer geographic proximity to company headquarters, changes in sentiments that could be associated with releases of earnings announcements, and the importance of chatter to millennial investors versus formal research that large financial institutions rely on to drive informed trades.

6. References

"AAPL." *Quiver Quantitative.*

"Amazon.com, Inc. (AMZN) Stock Historical Prices & Data." *Yahoo! Finance*, Yahoo!.

"AMC Entertainment Holdings, Inc. (AMC) Stock Historical Prices & Data." *Yahoo! Finance*, Yahoo!.

"AMC." *Quiver Quantitative.*

"Apple, Inc. (AAPL) Stock Historical Prices & Data." *Yahoo! Finance*, Yahoo!.

"AMZN." *Quiver Quantitative.*

Barber, Brad M. and Huang, Xing and Odean, Terrance and Schwarz, Christopher. "Attention Induced Trading and Returns: Evidence from Robinhood Users." *Journal of Finance, Forthcoming*, 12 October 2021. Available at SSRN: https://ssrn.com/abstract=3715077

Diangson, Bryan and Jung, Nicholas. "Bet it on Reddit: The Effects of Reddit Chatter on Highly Shorted Stocks." *Pomona College*, 1 May 2021.

Eaton, Gregory W. and Green, T. Clifton and Roseman, Brian and Wu, Yanbin. "Retail Trader Sophistication and Stock Market Quality: Evidence from Brokerage Outages." 15 October 2021. Available at SSRN: https://ssrn.com/abstract=3776874

Fitzgerald, Maggie, and Ari Levy. "Robinhood Valued at $32 Billion after Selling Shares in IPO at $38 per Share." *CNBC*, CNBC, 28 July 2021.

"FB." *Quiver Quantitative.*

"GameStop Corp. (GME) Stock Historical Prices & Data." *Yahoo! Finance*, Yahoo!.

"GME." *Quiver Quantitative.*

Heath, Thomas. "Shares by the Slice: Fractional Investing Sparks a Stock Market Stampede." *The Washington Post*, WP Company, 10 July 2020.

Hu, Danqi and Jones, Charles M. and Zhang, Valerie and Zhang, Xiaoyan. "The Rise of Reddit: How Social Media Affects Retail Investors and Short-sellers' Roles in Price Discovery." 14 March 2021. Available at SSRN: https://ssrn.com/abstract=3807655

Ingram, David, and Jason Abbruzzese. "Robinhood and Reddit: A Timeline of Two Apps Tormenting Wall Street." *NBCNews.com*, NBCUniversal News Group, 28 Jan. 2021.

Kaniel, Ron and Saar, Gideon and Titman, Sheridan. "Individual Investor Trading and Stock Returns." The Journal of Finance, 63: 273-310, 10 January 2008. Available at SSRN: https://doi.org/10.1111/j.1540-6261.2008.01316.x

Long, Cheng and Lucey, Brian M. and Yarovaya, Larisa. "'I Just Like the Stock' versus 'Fear and Loathing on Main Street': The Role of Reddit Sentiment in the GameStop Short Squeeze." 8 April 2021. Available at SSRN: https://ssrn.com/abstract=3822315

"Meta Platforms, Inc. (FB) Stock Historical Prices & Data." *Yahoo! Finance*, Yahoo!,

"Microsoft Corporation (MSFT) Stock Historical Prices & Data." *Yahoo! Finance*, Yahoo!.

"MSFT." *Quiver Quantitative*.

Primozic, Casey and Alex J. *Robintrack*.

"Tesla, Inc. (TSLA) Stock Historical Prices & Data." *Yahoo! Finance*, Yahoo!,

"TSLA." *Quiver Quantitative*.

7. Appendix

Table A1: Descriptive Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
ClosingPrice	5,733	473.025	865.103	1.98	3,731.41
RHUsers	3,474	146,852.35	118,895.44	4,924	718,187
TotalVolume	5,733	39,500,728	59,000,972	546,000	1.222e+09
WSBMentions	5,232	185.297	833.232	1	19,446
TwitterFollowers	1,713	7,525,907.3	3,731,784.4	6,548	13,494,773

Table A2: Summary of Closing Price

Stock Ticker	Mean	Std. Dev.	Freq.
AAPL	86.75	37.20	819
AMC	14.00	12.46	819
AMZN	2,452.37	713.17	819
FB	229.59	66.12	819
GME	48.98	75.77	819
MSFT	180.99	60.85	819
TSLA	298.48	286.58	819
Total	473.03	865.10	5,733

Table A3: Summary of WSB Mentions

Stock Ticker	Mean	Std. Dev.	Freq.
AAPL	110.63	198.83	817
AMC	320.51	1,447.68	575
AMZN	63.19	95.05	817
FB	48.81	100.38	813
GME	564.81	1,905.63	578
MSFT	79.74	139.23	817
TSLA	259.97	400.43	815
Total	185.29	833.23	5,232

Table A4: Summary of Robinhood Users

Stock Ticker	Mean	Std. Dev.	Freq.
AAPL	259,350.48	103,865.20	500
AMC	29,675.79	43,607.06	500
AMZN	137,622.35	69,341.55	492
FB	163,409.45	30,897.44	501
GME	16,103.65	9,870.36	493
MSFT	257,700.02	121,215.07	494
TSLA	163,624.86	96,844.02	494
Total	146,852.35	118,895.44	3,474