Is there good investment advice on r/wallstreetbets?

William Wildi

Supervised by : **Xiaowen Dong**

Thesis submitted for the degree of *Master of Science*

April 27, 2024

Is there good investment advice on r/wallstreetbets?

William Wildi

Acknowledgements

I would like to acknowledge my collaborators on this project: Valentina Semenova and Dragos Gorduza, who have helped with data preparation, analysis and were invaluable in our discussions about how to drive this work forward. I am grateful for the consistent help and support of my supervisors Xiaowen Dong and Stefan Zohren. Julian Winkler also helped by sharing some of his code used to process our Reddit data. Thank you to AlphaVantage for providing the stock price data used in this thesis.

Is there good investment advice on r/wallstreetbets?

William Wildi

Abstract

r/wallstreetbets (WSB), a subreddit where members informally share investment advice, garnered widespread interest due to its involvement in the 2021 Gamestop (GME) short squeeze. Here, we analyse whether posts on WSB can be used to predict stock returns. Looking at both the overall sentiment of the forum, and a subset of posts identified as due diligence (DD) reports, we find the forum provides little to no signal about future asset prices (with the exception of 'meme' stocks like GME and AMC). Given the forum's focus on high risk positions, retail investors are at risk of substantial losses if they follow the advice or sentiment of WSB.

Contents

CHAPTER 1

Introduction

Financial news made front page headlines in January 2021, but not for the typical reasons of a market crash or inflationary worries. News outlets were instead writing about video game retailer Gamestop (GME), whose share price rose from $17.35 on January 5, to a peak of $483 on January 28 – approximately a 28-fold increase.

Prior to this meteoric rise, GME was the most heavily shorted stock on the New York Stock Exchange [1], with short positions taken by institutional investors such as Melvin Capital. The price surge was triggered by users on a subreddit called r/wallstreetbets (WSB), an internet forum consisting largely of amateur retail investors. Throughout January, retail investors bought GME shares heavily, leading to a short squeeze and further price increases as many institutional investors exited their short positions to stem heavy losses.

WSB appears to have also driven extraordinary price volatility for other stocks in 2021, such as AMC Theatres, Blackberry and Nokia, leading to the term 'meme stock' entering the popular lexicon. These examples make it clear that retail traders are capable of significantly influencing stock markets, making this an important area to study for both finance researchers and institutional investors.

Forums like WSB are often seen as an obvious challenge to regulators: partly because the collective behaviour of the forum is arguably a form of market manipulation, and partly because of concerns that individual retail investors risk substantial losses by following WSB advice [2]. Implicit in the latter criticism is an assumption that WSB authors are unsophisticated noise traders – trading on information that is either irrelevant, or already incorporated in the price. This may be a mistake. Retail investors in the past had little more information to rely on than their own research, and the opinions of columnists and TV personalities; today, they are much more able to collate and share information with other retail traders online. This change in technology could imply a change in the behaviour and impact of retail investors on stock prices.

The objective of this paper is to assess whether posts on WSB are predictive of changes in asset prices. Broadly, there are three non-exclusive mechanisms that could underpin such a relationship. The first, and most obvious, is that retail investors can coordinate in sufficient number to actually move market prices in some scenarios (e.g. the GME and AMC short squeezes). This phenomenon is new, but not controversial, and there is already evidence that social media sentiment can be predictive of asset prices for 'meme' stocks like GME [3, 4].

The second way WSB could be predictive of market prices is through a 'wisdom of the crowd' type mechanism – where the aggregated advice of the forum turns out to be predictive of future asset prices. We assess this mechanism by comparing the relationship between the sentiment of the forum and asset prices.

The third possible mechanism is that a small subset of WSB members are sharing price-relevant information about particular assets, and this information is unknown or neglected by the larger market. While identifying this subset is challenging, posts on the forum containing more detailed analysis are tagged on the forum as 'DD' (due diligence).

DD posts typically express a clear prediction about a future asset price, along with reasons supporting positions, and thus are plausibly the best candidates for finding relevant information.

Our key finding is that WSB activity does not generally provide a useful signal for predicting future returns, except in the case of meme stocks (e.g. GME and AMC), where trading activity driven by WSB caused large price fluctuations. Forum sentiment on particular stocks tends to be reactive, i.e. large price changes precede forum activity, but it fails to provide a signal about future returns. When we look at DD posts specifically, we find that the recommendations of WSB authors are not associated with future returns in general. The only exception to this is in 2021, where DD recommendations were negatively associated with future returns, suggesting that the advice is at best noise, and at worst following it might lead to worse portfolio performance than performing random trades.

CHAPTER 2

Related Work

2.1 Social Media and Finance

Our paper takes inspiration from an emerging literature looking at the relationship between social media activity and stock market prices. It is well understood that new information about a company (such as from a news report) can affect its stock price, as well as the stock price of similar or competing companies [5]. Social media data provides information beyond news: shedding light on how individuals engage with financial information, which can reflect the sentiment of the market more broadly. Many researchers have exploited this to find associations between publicly expressed sentiment on Twitter and stock market prices [6–9]. Social media has thereby provided rich new text data for quantitative finance researchers to exploit, and is now a key input (along with news articles and company reports) for NLP methods in finance (see [10] for a review).

The internet has also increased the capacity of retail investors to gather information, share it with others, and respond to the analysis of other investors. Instead of relying on weekly newspaper columns, retail investors can go to specialised websites like Seeking Alpha, whose content is crowd-sourced from individual investors. Under the theory of efficient markets, asset prices should fully reflect all available information [11], and therefore posts written on a site like Seeking Alpha should be independent of future returns.

Empirical research, however, is not so clear cut. A 2014 analysis of Seeking Alpha, for example, found that views expressed on the platform were predictive of future stock returns [12]. This finding does not necessarily imply that Seeking Alpha articles provide 'good investment advice' – even when retail investors are making decisions based on noise, asset prices can diverge significantly from fundamental values for long periods of time [13, 14]. The 2014 article acknowledges this, noting that out-performance could be driven by herding effects, however, they do find that the Seeking Alpha articles are also predictive of earnings surprises, which fits less cleanly with a noise trader explanation [12]. Supporting this finding, another paper [15] found that Seeking Alpha posts encourage retail trading, and the the resulting order imbalances are predictive of price changes. The authors similarly conclude that these price changes are most likely being driven by retail investors inferring useful information from the reports.

On the flip-side, other research has suggested that social media can mislead retail investors. Retail investors can potentially end up in 'echo-chambers', primarily searching for content that supports their existing views, and making investment choices from a place of cognitive bias [16]. Social media also provides opportunities for rumours to spread (for example, about a potential merger), leading to uninformed trading behaviour and inhibiting price discovery [17]. Other papers have failed to find a relationship between investment opinions shared on social media and market variables [18]

Ultimately, whether social media is empowering or inhibiting retail traders is a question that is unlikely to be resolved, since the relationship between retail investor activity on social networks and market prices is likely to be sensitive to the data source used, the companies that are assessed, the time period, and the methodology of the researchers.

2.2 r/wallstreetbets

r/wallstreetbets (WSB) is a subreddit created in 2012 to discuss share and option trading. The forum particularly encourages aggressive, high risk trading strategies, and individuals frequently share information about large profits and losses resulting from the bets made [19]. The forum received limited academic interest until the GME short squeeze in early 2021 made WSB widely familiar. It is now one of the top 100 largest subreddits by subscriber count, with over 15 million users as of 2024 [20].

Most papers written on the forum have focused on the GME short-squeeze (and sometimes other 'meme' stocks) – either looking quantitatively at the effect WSB had on the GME stock price, or more qualitatively, trying to understand the emergence of WSB as a social phenomenon [3, 4, 21, 22]. Few papers have looked at the forum more broadly, one exception is [23], which provided a qualitiative assessment of how the WSB community has developed. Winkler and Semenova [24] were some of the first authors to look at the behaviour of WSB members in relation to retail trading, analysing the effect of sentiment contagion on the forum and its effect on increasing trading volume.

More recent papers have attempted to assess if WSB posts can be used to predict stock price changes. One of the first papers on this subject looked at the returns of the top 20 stocks mentioned on WSB, and found that an equally weighted portfolio of these stocks significantly outperformed the S&P 500 [25]. This finding should be treated with caution, as the authors based the portfolio on the most popular stocks from 2018-2021, and used the same time period to measure performance, meaning that the forum may simply prefer stocks that have performed well (there is also no accounting of volatility). The authors also identified 'buy' signals from the forum by doing a word count of 'buy' and 'sell' across posts, and found that a 'buy' signal from the forum preceded stock price increases that were higher than the S&P 500, and also higher than days where no buy signal was found. However, the authors again appeared not to risk-adjust the returns in any way, and found significantly less impressive returns prior to 2021. Later work from other researchers has found WSB sentiment impacts volume and volatility, but it is not a useful price signal [26, 27].

The other major paper looking at WSB and returns takes a different approach: focusing on posts marked as ‘due diligence’ (DD) [28]. DD posts are supposed to contain market research, and typically also provide a prediction about how a future asset price is going to behave in the future. The authors found that from 2018-2020, DD recommendations are significant predictors of one-month ahead returns, and comments are incrementally useful when they agree with the sentiment expressed by the post author. However, from 2021 on, the informativeness of DD posts disappears, which the authors infer as evidence that the influx of users after the Gamestop short squeeze have decreased the quality of advice on the platform.

CHAPTER 3

Methods

Our primary research question is whether WSB submissions mentioning company i at time t can be used to predict the future returns of i at some point in the future $t + k$. We assess this both by looking at aggregate sentiment on the forum as an input, as well as the specific subset of due diligence (DD) reports on the forum.

3.1 Data

Submission and comment data from WSB is collected through the Pushshift Reddit database [29], which provides an archive of all posts and comments on Reddit. We collect data from the inception of the forum in April 2012 up to 24 June 2021. Posts on WSB typically refer to companies by their stock symbol, and we use the work of [24] to obtain a list of stock symbols (from Yahoo Finance and Compustat) and extract these from the submission text (with upper-case matching). For symbols that form common words/acronyms (e.g. WISH), we need to distinguish the use of the word written in upper-case from an explicitly reference to the stock. Fortunately, one of the communication norms of the forum is that tickers are supposed to be preceded with a dollar sign, so we only tag stock mentions of common words when preceded by the dollar sign. We inspected several thousand posts to ignore symbols referring to companies that are virtually never discussed on the forum, but whose symbol is used for other references. For example, the appearance of the letters SI almost always refers to short-interest, and never to the associated company: Silvergate Capital Corp. We also ignore any reference to SPY (S&P 500), since we want to use this as a baseline.

For simplicity, we only keep companies traded on the NYSE or NASDAQ, since this allows us to operate in a consistent timezone. This is also a fairly innocuous step, since the vast majority of WSB activity is focused on the US and companies that trade in US markets. After all these steps are taken, we filter our data to only include posts that mention exactly one stock.

Stock price data comes from AlphaVantage. AlphaVantage helpfully provides an adjusted close price to take into account events like dividends and stock splits, which we use to calculate returns. We collect a daily time series for each stock in our data set, and we remove posts when AlphaVantage did not have data on a particular stock. At the end of this process, we ended up with 202,681 submissions, with 6,320,968 corresponding comments. As forum activity has increased by several orders of magnitude since inception (Figure 3.1.1), our data is concentrated in more recent months, with the 2021 data accounting for over 60 percent of posts, despite only accounting for around 5 percent of the forum's lifetime.

To obtain the subset of due diligence (DD) reports in the forum, we firstly filter to posts that have the DD flair. Flairs are used on subreddits to categorise posts, and posts that are flaired as DD are supposed to be moderated by WSB admins. This is the same method used by [28]. We note, however, two issues with this approach: firstly, the DD flair does not appear to be used in any significant way until July 2018, so this excludes

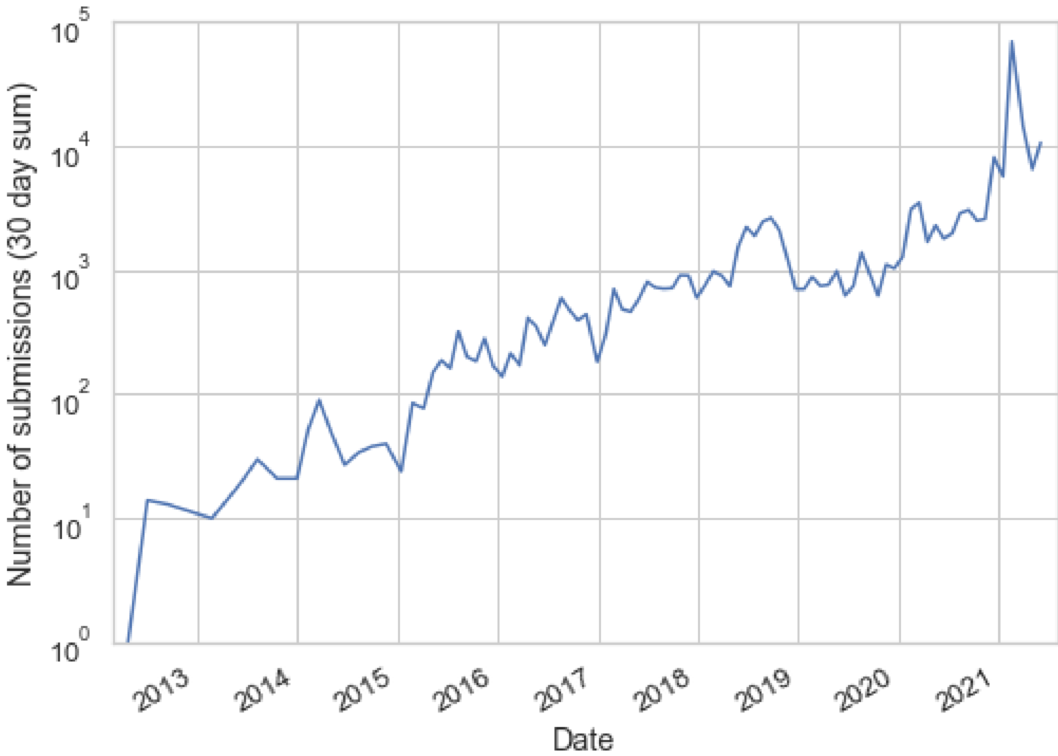


Figure 3.1.1: Total number of new posts on WSB every 30 days

any older posts. Secondly, we noticed upon reading the DD posts that quality control is limited in our dataset. The flair guidelines state that DD posts must be a high-effort, longer, researched posts, but many posts in the data set fail to meet this criteria. This is partially because Pushshift retains many posts that are later removed by moderators, but many posts that were not removed by moderators still do not appear to meet the criteria.

To resolve these issues, we take two additional filtering steps beyond [28]. Firstly, we supplement the posts containing the DD flair with posts that contain the DD acronym. Secondly, we manually review every single potential DD post (i.e. flaired posts that are not removed by moderators, or contain the text 'DD'), and we remove any post that does not appear to be a valid DD. For us to consider a post a valid DD, we require:

1. The post is at least 20 words long
2. The post is primarily about one company
3. The post has a positive or negative sentiment on the company's future price

4. The post attempts to provide a reason for the future price to change (e.g. sharing recent news about the company, providing analysis of recent trends or state of the market etc.)

The sentiment of the post is determined by looking for clear indicators that the post author believes the price is going to increase or decrease. In most cases, this is not too difficult to determine, since one of the norms of WSB is that posters are supposed to state their trading positions, which could be buying/shorting assets, or purchasing/selling options.

Inevitably, there is some subjectivity in this filtering, especially for the fourth criterion. We attempt to be quite permissive in assessing these criteria; for example, we would allow news about a new product release to be used as a reason (even though this is very likely to already be incorporated into the price). We do not, however, permit reasons that are excessively vague; for example, saying a company's price will increase because 'fundamentals are good' is insufficient without some added specificity of what is good about the fundamentals. These filtering criteria remove around 40 percent of our candidate posts, leaving us with 3,650 DD posts.

3.2 Sentiment detection

The DD posts are sufficiently small in number that we can manually label each post as bullish/positive or bearish/negative manually (neutral posts are removed). We find 77 percent of posts are bullish, and 23 percent are bearish. For the forum as a whole, however, there are far too many posts for manual classification. Previous papers have either relied on counting keywords (for example, 'buy', 'sell', 'call', 'put' etc. [25, 28]), or they have used off-the-shelf rule based sentiment classifiers [4], such as VADER [30].

The advantage of such approaches is their simplicity, but they have their problems. Choosing keywords/phrases is inherently arbitrary – it could easily miss posts that are clearly positive/negative because they don't contain the required keywords, and it is vulnerable to mistakes when authors negate phrases in the dictionary (e.g. 'calls are going to be worthless'). Algorithms like VADER by contrast lack the specific domain vocabulary of the WSB forum, hence their performance can be quite poor [4].

Another issue with using VADER is that we are not interested in the general sentiment of WSB posts *per se*, instead, we are interested in whether the poster has a bullish, bearish, or neutral attitude towards the future price of an asset. To clarify, suppose a trader talks about how much money they made from a recent trade on Tesla; the post is likely to contain positive sentiment towards Tesla, but it is backwards looking, and the author may take no position on the future value of Tesla shares. We want to categorise such a post as neutral, despite it having an overall positive valence.

Rather than rely on rule-based sentiment detection, we opt for using a pre-trained, fine-tuned NLP model based on transformer architecture [31]. We chose to use the FinBERT model [32], which itself was developed from the original pre-trained BERT model. FinBERT was trained in two stages: firstly, the model was trained on financial news data from Reuters to familiarise the model with domain specific language. Secondly, the model was then fine-tuned to perform sentiment detection on a financial phrasebank developed in a previous research project [33].

While FinBERT's training will ensure that it has familiarity with finance jargon, our use-case clearly has some important differences to FinBERT's training environment. Not only are we trying to predict a more specific kind of sentiment (as aforementioned), but the language of WSB certainly has clear differences with finance news, even though many of the specific terms will be similar. To resolve these issues, we fine-tuned FinBERT on our own data. We manually coded a random sample of 4,000 WSB posts, categorising them as bullish, bearish, or neutral depending on what the author's attitude is on future price increases. Of the posts in the sample 41 percent are bullish, 37 percent are neutral and 22 percent are bearish. [1]

[1]Since bearish posts tend to be a minority on the forum, bearish posts were up-sampled to reduce class imbalance.

The sample was split, with 75 percent of data used for training, 10 percent for validation and 15 percent for testing. The maximum text length was 256 words (shorter posts were padded, longer posts were truncated), and we ran batch sizes of 8[2]. Validation accuracy was assessed every 100 batches, and training was stopped when cross-validation accuracy failed to increase after 10 evaluations.

Our final model had a test accuracy of 69.2 percent. While this may not seem especially accurate, it is important to note that the unique language of the forum combined with our requirement of predicting future-facing sentiment made this task challenging. When a sample of the manually annotated data was checked by a second human, we found it was not possible to unambiguously assign a sentiment label in around 10-15 percent of cases. These ambiguous cases were not removed from the training data, since we wanted to make sure the model received a representative training sample of the forum, but this makes producing a highly accurate model very challenging.

Nonetheless, we believe the performance is satisfactory for our purposes. It is substantially better than a majority class baseline (41 percent), and it is a significant improvement over [24], whose model only achieved a test accuracy of 57.5 percent. The confusion matrix is shown in Table **??**. Note that relatively few errors are caused by misclassifying bullish posts as bearish (or vice versa) – this is a good sanity check of the model, as we would expect differentiating between bullish/neutral or bearish/neutral to be more difficult than distinguishing bullish/bearish posts from each other.

Table 3.1: Confusion matrix showing the results of the fine-tuned FinBERT classifier on the test data ($n = 604$)

		Prediction		
		Bullish	Bearish	Neutral
Label	Bullish	173 (28.6%)	30 (5.0%)	49 (8.1%)
	Bearish	14 (2.3%)	95 (15.7%)	21 (3.5%)
	Neutral	42 (7.0%)	30 (5.0%)	150 (24.8%)

[2]Fine tuning was performed on a Nvidia RTX 3070 GPU, which ran into memory issues when larger batch sizes were attempted

3.3 Stock prices and WSB sentiment

We investigate the relationship between aggregate sentiment in two ways. Firstly, we want to look at how asset prices are changing shortly before and after posts on the forum. To do this, we follow the methodology of [5], who looked at market movements around news events by looking at changes in abnormal return (AR). AR is derived from the Capital Asset Pricing Model [34], which models returns for company i at time t as follows:

$$r_{i,t} = \alpha_t + \beta_t r_{m,t} + \epsilon_{i,t}, \tag{3.1}$$

where $r_{i,t}$ is the log return in the price of stock i on day t compared with the previous day, i.e. $r_{i,t} = \log(\frac{p_{i,t}}{p_{i,t-1}})$ where $p_{i,t}$ is the (adjusted close) stock price. $r_{m,t}$ is the return of the market (in our case, we use the S&P 500), hence β captures stock price moves that are driven by movements in the wider market. $\alpha_{i,t}$ captures stock over/under-performance relative to the market, and $\epsilon_{i,t}$ is a stochastic error term, often referred to as abnormal return (AR).

AR tends to have high magnitude in the presence of sudden price shocks (for example, a news story about a particular company), and hence it is often used for event detection in financial markets. In our case, it is useful for assessing if WSB sentiment is able to provide indication about a future price shock. The CAPM model is fit for a given company i on day t by looking at the daily returns of i over the previous 180 days, and fitting an OLS model. Following [5], we calculate the seven-day cumulative abnormal retun (CAR).

$$CAR_{i,t} = \sum_{t-6}^{t} \epsilon_{i,t}. \tag{3.2}$$

Once we round the timestamp of a WSB submission to the nearest future market close time, we can match the CAR for each company/time to the WSB data, along with a time series of how the CAR changed over the previous 14 trading days, and the following 14 trading days. We can then group the posts by stock and sentiment to produce our CAR plots (see Figure 4.1.1).

We extend this analysis further by looking at whether sentiment is useful for forecasting future returns by conducting a Granger causality test [35]. To get a time-series of sentiment, we firstly need a sentiment score for each company and date. Let $s_{i,t}^{(j)}$ represent the net sentiment of post j, which is calculated by taking the (softmax) output of our sentiment classifier, and calculating the difference between the bullish and bearish outputs (hence $-1 \leq s_{i,t}^{(j)} \leq 1$). We can then say the net sentiment of each stock i on day t is:

$$S_{i,t} = \sum_{j} s_{i,t}^{(j)}. \tag{3.3}$$

$S_{i,t}$, however, has a tendency to increase over time, since the forum grows in popularity, and most posts are bullish. To resolve this, we normalise $S_{i,t}$ by dividing it by the average number of posts on the forum over the past seven days (call this n_t). We then calculate the difference in normalised sentiment to get a value for our sentiment time series $\Delta \hat{S}_{i,t}$ where:

$$\Delta \hat{S}_{i,t} = \hat{S}_{i,t} - \hat{S}_{i,t-1} = \frac{S_{i,t}}{n_t} - \frac{S_{i,t-1}}{n_{t-1}}. \tag{3.4}$$

3.4 Due diligence reports

3.4.1 Variables

Since DD reports only form a small fraction of total WSB activity, the methods used for investigating the relationship between average sentiment and future returns are not appropriate, as these relied on averaging over thousands of data points per stock, and constructing time series for the Granger causality tests. Instead, it is better to consider whether the event of each DD post precedes a change in return. We largely follow the methods of [28] in our analysis of the DD reports.

Our main variable of interest is the sentiment of posts on day t for company i, which we will denote $d_{i,t}$. This is very similar to our definition of $S_{i,t}$ from the previous section, except is a binary label instead of a softmax output. For the analysis of the DD posts, we separate bullish posts from bearish posts and fit models on each of them instead of computing a net score. The justification for this choice is that we wanted to consider the effect of variables beyond sentiment, and this is much more straightforward when the bullish/bearish posts are treated separately.

Beyond sentiment, we defined a series of additional variables that might be a relevant indicator of the quality of a post:

1. The number of comments a post receives. We consider the submission to be a comment so each submission has at least one comment.
2. The number of words the post contains.
3. The maximum depth of a submission comment. We consider the submission itself to have a depth of zero, comments replying to the submission have a depth of one, comments replying to comments replying to the submission have a depth of two and so on. This value is divided by the log number of comments in the post, otherwise it is highly correlated with our first variable.
4. A binary indicator whether a post contains a URL. URLs could be to another Reddit post or an external source, but urls containing images were filtered out.
5. A binary indicator whether a post was 'proactive'. We consider a post to be proactive if the post occurred when the CAR of the stock was not statistically significantly different from zero (using a t-test). Contrast this to 'reactive' posts, which are talking about companies whose price is substantially different from what we would expect under the CAPM.

All variables were aggregated by summing across each unique day/company combination, and the first two variables were also log transformed, since these variables have very heavy tails. Since two of our variables depend on actions taken after the post is submitted, we only include comments that occur within 24 hours of the original post.

The choice of each of these features was motivated by various considerations. Number of comments and maximum depth can both be considered as graph features when the Reddit submission and successive comments are represented as a directed acyclic graph, also known as an information cascade. Previous research on Twitter has found that graph features can be predictive of the content of tweets, which has been a particularly active area of research in fake news detection [36–38]. If Reddit users react differently to posts with more predictive power (perhaps because they are better analysed), then such graph features could be useful for our models.

The other three features serve as proxies for the quality of a post. Longer posts may have more detailed robust analysis, and therefore may be more predictive of price changes. Posts that contain URLs are providing sources, which again could be an indicator of a more reliable signal. The proactive feature is inspired by one of the previous WSB papers [25], which suggested that when a submission is created while the price of an asset has not changed recently, then it might be more likely to contain a useful signal compared to posts about assets whose price is already very far from its typical value (and therefore likely to experience mean-reversion).

3.4.2 Models

For each day and company in our dataset, we calculate a net DD sentiment score. We then fit the following fixed effects model:

$$r_{i,[t,t+m]} = \beta_1 d_{i,t} + \beta_2 x_{i,t,k} + \beta_3 r_{i,[t,t-1]} + \beta_4 CAR_{i,t} + \gamma_t + \epsilon_{i,t}, \tag{3.5}$$

where $r_{i,[t,t+m]}$ is the (log) return between day t and day $t+m$, $d_{i,t}$ represents our sentiment variable, and $x_{i,t,k}$ is one of our additional variables, where k denotes which variable is being used. $r_{i,[t-1,t]}$ is the return of the stock when comparing the previous day price to day t, and γ_t is our fixed time effects. Time fixed effects ensure that we control for overall movements of the markets, but including fixed effects for firms is inappropriate for two reasons. Firstly, it would mean that our model was including future information about a company's performance, and secondly, if the forum is good at picking stocks that generally outperform the market (even if they do not get the timing especially precise), then we want to be able to observe that signal in our model. To estimate our fixed effects and control variables accurately, we include closing price data for all stocks mentioned in the forum at least once, and include prices from 2017 July 17 up to 21 June 2021, which is the date range of our DD sample. This gives each model around 1.5 million data points to estimate our time effects and control variables.

The use of log-returns is motivated by the fact that log-returns are symmetric for increases and decreases, while arithmetic returns are not. For example, if an investment of £100 decreases by x percent in one period, then increases by x in the following period, at the end of the second period, the investment will then be worth $100(1-x^2) < 100$ if $x > 0$. If, however, an investment has a log-return of $-x$ in the first period, and x in the second, then the investment will still be worth £100. This symmetry avoids the problem of the model under-fitting on assets that underwent a substantial loss, since arithmetic returns can be arbitrarily large on the positive side, but cannot go lower than negative one. It also more realistically reflects how an asset manager would make investment decisions to maximise long-term wealth [39].

Since some of our variables depend on comment activity after the initial post, and we truncate all comments that occur after 24 hours, the day t used to get the price of the asset used to calculate $r_{i,[t,m]}$ is defined as the first open or close price 24 hours after the post was created. This is done to block reverse causality, since it is possible that number of comments or maximum comment depth increase after price changes, which would show as a misleading association in our model.

Having previous day return and CAR included in our model has a couple of benefits. Firstly, depending on the size of m, the control variables can help to control for price changes resulting from momentum or mean reversion [40]. Second, as we saw in our previous section, forum activity tends to be highest when CAR is very high, so including these controls ensures that any effect measured is not simply a proxy for prices already being high. To ensure our control variables are estimated precisely, our model includes all companies and trading days from the creation date of the earliest DD post, up to the creation date of the last DD post.

We fit 3,744 models, the number resulting from the multiplication of the following configurations:

- Bullish and bearish posts are considered in separate models – this makes aggregation of the remaining independent variables considerably more straightforward (2 configurations).
- Values of m are considered from one to 26 weeks in one week intervals (26 configurations).
- Every model has sentiment as an independent variable, and zero or one of our other independent variables (6 configurations).
- Whether the model considers all posts, or just flaired posts (2 configurations).
- Whether the model considers all stocks, or excludes meme stocks: GME, AMC, BB, and NOK (2 configurations). We select these four assets because their stock price is commonly considered to have been impacted by WSB trading activity [3]
- Whether the model considers all posts, posts before 2021, and posts from 2021 (3 configurations).

Having multiple configurations allows us to discern whether any effects we observe are robust to alternative configurations. We therefore not only consider whether a parameter is statistically significant in a single model, but whether it is statistically significant across multiple considerations.

[3] We also find evidence for this in our Granger causality tests discussed in the results section.

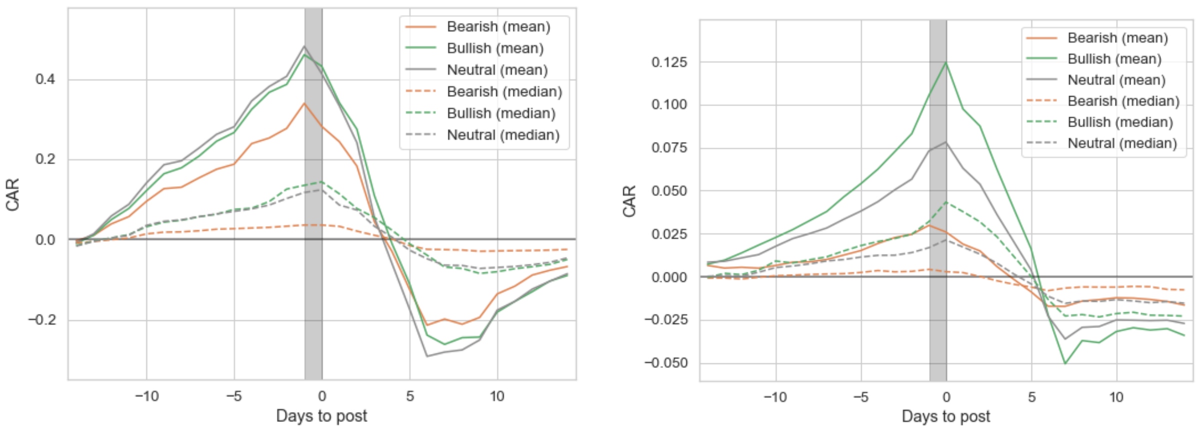
CHAPTER 4

Results

4.1 WSB sentiment and returns

Figure 4.1.1 shows how average cumulative abnormal return (CAR) changes 14 trading days before and after a submission. Figure 4.1.1(a) shows the CAR plots for all data, while Figure 4.1.1(b), excludes GME and AMC, as the short squeeze events on these stocks created exceptionally high abnormal returns. We observe the following: firstly, CAR tends to be rising up to the submission date, and rapidly declines afterwards. This suggests that WSB activity tends to follow significant price changes in the market, rather than providing a leading indicator of price changes. Secondly, the shape of the curves are quite similar, regardless of sentiment breakdown, and it is primarily the magnitude of the CAR that differs. This suggests that the greater the CAR of a particular asset, the more likely it is that the asset will be discussed on the forum across all sentiment classes, although the effect is more pronounced for bullish posts.

One might attempt to infer from the sharp fall in CAR after posts that shorting stocks that are popular on WSB might be a profitable trading strategy. This inference is, however, mistaken for two reasons. Firstly, negative CAR does not necessarily imply negative returns. Secondly, and more importantly, the CAR plots suggest that WSB sentiment peaks when CAR peaks, which means that predicting when sentiment will peak is not straightforward to predict. If sentiment is already high, and AR increases, then sentiment is likely to continue to increase.



(a) (b)

Figure 4.1.1: Average cumulative abnormal return 14 trading days before and after post submission, grouped by post sentiment. The black bar shows the window of time when posts were written. (a) uses all posts in the data, while (b) excludes GME and AMC.

Figure 4.1.2 shows a breakdown of the CAR plots for the twenty most popular stocks on the forum[1]. We see similar patterns to Figure 4.1.1 for smaller 'memey' stocks like GME, AMC, BB, SPCE, and NOK. This pattern can also be observed for TSLA and AMD. Other large companies like AAPL, MSFT, AMZN, NVDA, GE and BABA are fairly flat in their pattern, suggesting a weaker relationship between sentiment and returns. The only two cases where sentiment does seem to meaningfully precede peaks in CAR are SNDL and SNAP. The performance of SNDL (a cannabis company) appears by a spike in popularity in late January 2021, which preceded a tripling in the share price over the next few weeks. The performance of SNAP also appears to be driven by outliers, as its median CAR chart is quite flat compared to its mean. Overall, the CAR analysis

[1]Palantir Technologies (PLTR) and Rocket Companies (RKT) are more popular stocks on WSB than General Electric (GE) or Disney (DIS), however, both stocks had their IPOs in 2020, so there is not sufficient data to fit the CAPM model for these stocks.

suggests that besides a few outliers, WSB sentiment appears to be a lagging indicator of stock price changes, not a leading one.

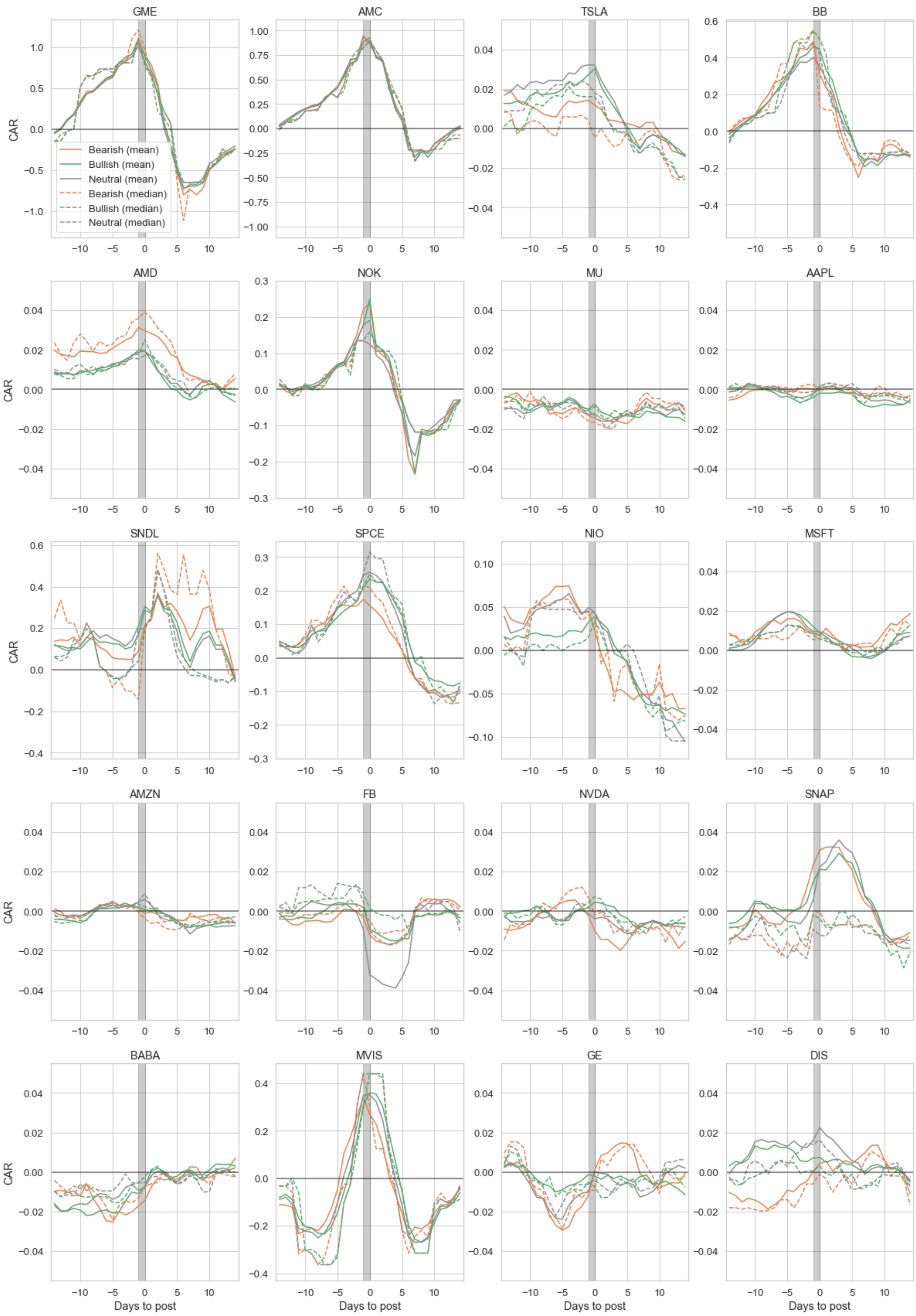


Figure 4.1.2: Average cumulative abnormal return 14 trading days before and after post submission, grouped by post sentiment for each of the 20 most popular stocks on WSB.

Next we check whether our normalised sentiment measure $\Delta \widehat{S}_{i,t}$ is Granger causal of daily stock price returns, i.e. we are assessing whether lagged values of WSB sentiment is useful for forecasting future returns. We use time-series for the top 22 most popular stocks on the forum,[2] starting from the earlier of 1 January 2016, or the first time a stock was mentioned on the forum. The removal of pre-2016 data is motivated due to the small size of the forum before this point, which can result in volatile changes in our sentiment measure.

For all 22 stocks, our sentiment measure and the daily log-return time series appear stationary, which we check with an augmented Dickey-Fuller test. We run Granger Causality tests at lags of 1, 2, 5, and 10 trading days. The results are shown in Table 4.1.

Table 4.1: Results of Granger causality tests for log-return for the top 22 most popular stocks on WSB. Main numbers show the Wald test statistic, with p values in parentheses. * are used to indicate significance levels.

	Obs	1	2	5	10
AAPL	1338	0.140	0.047	0.439	1.177
		(0.708)	(0.954)	(0.821)	(0.302)
AMC	1184	9.833	36.189	21.110	10.551
		(0.002)**	(0.000)***	(0.000)***	(0.000)***
AMD	1336	0.801	2.182	1.446	1.147
		(0.371)	(0.113)	(0.205)	(0.323)
AMZN	1334	0.224	0.183	1.735	2.033
		(0.636)	(0.833)	(0.123)	(0.027)*
BABA	1339	1.055	1.329	1.260	0.887
		(0.305)	(0.265)	(0.279)	(0.545)
BB	1212	31.792	14.402	16.236	12.083
		(0.000)***	(0.000)***	(0.000)***	(0.000)***
DIS	1326	0.410	0.760	0.731	0.950
		(0.522)	(0.468)	(0.6)	(0.486)
FB	1338	0.303	0.147	0.620	0.815
		(0.582)	(0.863)	(0.685)	(0.614)
GE	1301	0.329	0.901	1.118	1.037
		(0.566)	(0.406)	(0.349)	(0.409)
GME	1284	28.820	16.922	16.206	16.749
		(0.000)***	(0.000)***	(0.000)***	(0.000)***
MSFT	1325	2.313	1.176	0.928	0.816
		(0.129)	(0.309)	(0.462)	(0.614)
MU	1295	0.120	0.128	0.098	0.920
		(0.729)	(0.88)	(0.992)	(0.514)
MVIS	261	0.306	1.658	1.322	0.632
		(0.58)	(0.192)	(0.255)	(0.786)

Continued on next page

[2]For consistency with our CAR plots, we include GE and DIS, in addition to RKT and PLTR

Table 4.1 – continued from previous page

	Obs	1	2	5	10
NIO	637	0.169	0.645	0.240	0.704
		(0.682)	(0.525)	(0.945)	(0.721)
NOK	1319	10.168	5.884	3.476	2.003
		(0.001)***	(0.003)**	(0.004)**	(0.03)*
NVDA	1336	0.093	0.168	0.494	0.617
		(0.761)	(0.845)	(0.781)	(0.801)
PLTR	130	0.202	0.839	1.088	2.065
		(0.654)	(0.434)	(0.371)	(0.034)*
RKT	167	0.168	0.529	0.283	0.287
		(0.682)	(0.59)	(0.922)	(0.983)
SNAP	1022	13.782	6.907	2.815	1.811
		(0.000)***	(0.001)***	(0.016)*	(0.055)
SNDL	327	0.312	3.500	1.810	1.212
		(0.577)	(0.031)*	(0.111)	(0.283)
SPCE	391	3.901	1.169	1.298	2.017
		(0.049)*	(0.312)	(0.264)	(0.031)*
TSLA	1334	0.293	0.369	0.627	0.393
		(0.588)	(0.691)	(0.679)	(0.95)

The results show that for most stocks, WSB sentiment is not useful for forecasting returns. This is not true, however, for the meme stocks GME, AMC, NOK and BB, but this fits with the conventional wisdom that retail traders on WSB drove up the prices of these shares in early 2021.

We do get statistically significant results ($0.01 < p < 0.05$) for AMZN at a ten day lag, PLTR at a ten day lag, SNDL at a two day lag, and SPCE at a one and ten day lag. Given that sentiment is not Granger causal of returns at other lags for these stocks, this is most likely to be a spurious result.

The most interesting statistically significant results come from SNAP, where the test statistic is strongly significant at one and two days ($p < 0.001$), statistically significant at five days ($p < 0.05$), and almost significant at ten days ($p = 0.055$). This also coheres with the result of Figure 4.1.2. SNAP is not conventionally considered a meme stock, and posts mentioning it were much more common before 2021. This potentially suggests that WSB sentiment for SNAP may have been a more useful leading indicator in previous years.

4.2 Due diligence posts

The results of the fixed effects models for the DD posts are shown in Tables 4.2, 4.3, 4.4, and 4.5. Since we fitted over three thousand models, it is not practical to show all results, so instead presentation is limited to just the models looking at price changes over four weeks and twelve weeks, which are fairly representative of the results in general. The results for the models that only include the flaired DD posts are not included, since it was of little practical importance. Below we discuss the results for each of our independent variables.

Table 4.2: Panel regression results for the bullish post model where the dependent variable is the asset price change four weeks after the post. Standard errors are in parentheses, with * indicating levels of statistical significance.

Time	Meme Stocks	Other Variable Name	Sentiment	Other Variable	CAR	$r_{[t,t-1]}$
All posts	Included		-0.0102		-0.0356	-0.0048
			(0.0072)		(0.011)**	(0.0123)
		Num comments	-0.0345	0.0113	-0.0356	-0.0048
			(0.0134)**	(0.0079)	(0.011)**	(0.0123)
		Post word count	-0.0301	0.0052	-0.0356	-0.0048
			(0.0099)**	(0.0033)	(0.011)**	(0.0123)
		Max comment depth	-0.0316	0.0159	-0.0356	-0.0048
			(0.0232)	(0.0183)	(0.011)**	(0.0123)
		Proactive	-0.0404	0.0443	-0.0356	-0.0048
			(0.0121)***	(0.0119)***	(0.011)**	(0.0123)
		Includes URL	-0.0059	-0.0116	-0.0356	-0.0048
			(0.0089)	(0.0112)	(0.011)**	(0.0123)
	Exluded		-0.0112		-0.0368	-0.0049
			(0.006)		(0.0111)***	(0.0123)
		Num comments	-0.0106	-0.0002	-0.0368	-0.0049
			(0.0109)	(0.0038)	(0.0111)***	(0.0123)
		Post word count	0.0153	-0.0057	-0.0368	-0.0049
			(0.0158)	(0.003)	(0.0111)***	(0.0123)
		Max comment depth	-0.0063	-0.0036	-0.0368	-0.0049
			(0.0101)	(0.0057)	(0.0111)***	(0.0123)
		Proactive	-0.0389	0.0326	-0.0368	-0.0049
			(0.018)*	(0.0172)	(0.0111)***	(0.0123)
		Includes URL	-0.0094	-0.0049	-0.0368	-0.0049
			(0.0066)	(0.0094)	(0.0111)***	(0.0123)
Post-2021	Included		-0.0304		-0.0358	-0.005
			(0.0139)*		(0.011)**	(0.0123)
		Num comments	-0.0518	0.0144	-0.0358	-0.005
			(0.0107)***	(0.0116)	(0.011)**	(0.0123)
		Post word count	-0.0414	0.0041	-0.0358	-0.005
			(0.0055)***	(0.0045)	(0.011)**	(0.0123)
		Max comment depth	-0.0926	0.0495	-0.0358	-0.005
			(0.0319)**	(0.0297)	(0.011)**	(0.0123)
		Proactive	-0.0502	0.0463	-0.0358	-0.005
			(0.015)***	(0.0192)*	(0.011)**	(0.0123)
		Includes URL	-0.0205	-0.025	-0.0358	-0.005
			(0.0224)	(0.0294)	(0.011)**	(0.0123)
	Exluded		-0.0439		-0.0373	-0.005
			(0.0112)***		(0.0111)***	(0.0123)
		Num comments	-0.0456	0.0007	-0.0373	-0.005
			(0.0184)*	(0.0068)	(0.0111)***	(0.0123)
		Post word count	-0.0257	-0.0039	-0.0373	-0.005
			(0.0351)	(0.0061)	(0.0111)***	(0.0123)
		Max comment depth	-0.0515	0.0061	-0.0373	-0.005
			(0.0173)**	(0.0107)	(0.0111)***	(0.0123)

Continued on next page

		Proactive	-0.0807	0.0428	-0.0373	-0.005
			(0.0316)*	(0.0297)	(0.0111)***	(0.0123)
		Includes URL	-0.0477	0.0085	-0.0373	-0.005
			(0.0135)***	(0.017)	(0.0111)***	(0.0123)
Pre-2021	Included		0.0032		-0.0373	-0.005
			(0.009)		(0.0112)***	(0.0123)
		Num comments	-0.0019	0.002	-0.0373	-0.005
			(0.013)	(0.0063)	(0.0112)***	(0.0123)
		Post word count	0.0304	-0.0059	-0.0373	-0.005
			(0.0171)	(0.0035)	(0.0112)***	(0.0123)
		Max comment depth	0.0109	-0.0055	-0.0373	-0.005
			(0.0118)	(0.0082)	(0.0112)***	(0.0123)
		Proactive	-0.0158	0.0224	-0.0373	-0.005
			(0.023)	(0.0202)	(0.0112)***	(0.0123)
		Includes URL	0.0046	-0.0041	-0.0373	-0.005
			(0.0092)	(0.0107)	(0.0112)***	(0.0123)
	Exluded		-0.003		-0.037	-0.0052
			(0.0068)		(0.0111)***	(0.0123)
		Num comments	0.0045	-0.0029	-0.037	-0.0052
			(0.0121)	(0.0043)	(0.0111)***	(0.0123)
		Post word count	0.0251	-0.0061	-0.037	-0.0052
			(0.0181)	(0.0036)	(0.0111)***	(0.0123)
		Max comment depth	0.0091	-0.0087	-0.037	-0.0052
			(0.0117)	(0.0067)	(0.0111)***	(0.0123)
		Proactive	-0.0292	0.0309	-0.037	-0.0052
			(0.0205)	(0.0197)	(0.0111)***	(0.0123)
		Includes URL	-0.0012	-0.005	-0.037	-0.0052
			(0.0073)	(0.0111)	(0.0111)***	(0.0123)

Table 4.3: Panel regression results for the bearish post model where the dependent variable is the asset price change four weeks after the post. Standard errors are in parentheses, with * indicating levels of statistical significance.

Time	**Meme Stocks**	**Other Variable Name**	**Sentiment**	**Other Variable**	**CAR**	$\mathbf{r_{[t,t-1]}}$
All posts	False		-0.0037		-0.0363	-0.0029
			(0.0139)		(0.011)***	(0.012)
		Num comments	-0.0039	-0.0001	-0.0363	-0.0029
			(0.0204)	(0.0094)	(0.011)***	(0.012)
		Post word count	0.0234	-0.0063	-0.0363	-0.0029
			(0.0247)	(0.0063)	(0.011)***	(0.012)
		Max comment depth	-0.0008	-0.002	-0.0363	-0.0029
			(0.0245)	(0.0141)	(0.011)***	(0.012)
		Proactive	0.0211	-0.0331	-0.0363	-0.0029
			(0.0334)	(0.0363)	(0.011)***	(0.012)
		Includes URL	-0.012	0.0228	-0.0363	-0.0029
			(0.0162)	(0.0164)	(0.011)***	(0.012)
	True		0.0051		-0.037	-0.003
			(0.0125)		(0.0111)***	(0.0121)
		Num comments	-0.0076	-0.0046	-0.037	-0.003
			(0.0198)	(0.0081)	(0.0111)***	(0.0121)
		Post word count	0.0248	-0.0045	-0.037	-0.003
			(0.0228)	(0.0047)	(0.0111)***	(0.0121)
		Max comment depth	0.0049	0.0001	-0.037	-0.003
			(0.022)	(0.0136)	(0.0111)***	(0.0121)
		Proactive	0.0245	-0.025	-0.037	-0.003
			(0.0383)	(0.0378)	(0.0111)***	(0.0121)
		Includes URL	0.0004	0.0124	-0.037	-0.003
			(0.0145)	(0.0168)	(0.0111)***	(0.0121)
Post-2021	False		-0.0451		-0.0367	-0.0055
			(0.0354)		(0.011)***	(0.0123)
		Num comments	-0.0588	-0.0055	-0.0367	-0.0055
			(0.1576)	(0.0596)	(0.011)***	(0.0123)
		Post word count	0.029	-0.0197	-0.0367	-0.0055
			(0.1315)	(0.0342)	(0.011)***	(0.0123)
		Max comment depth	-0.2937	0.1583	-0.0367	-0.0055
			(0.2155)	(0.1456)	(0.011)***	(0.0123)
		Proactive	-0.0293	-0.0385	-0.0367	-0.0055
			(0.0721)	(0.1319)	(0.011)***	(0.0123)
		Includes URL	-0.0395	-0.0171	-0.0367	-0.0055

Continued on next page

			(0.0494)	(0.105)	(0.011)***	(0.0123)
	True		-0.0389		-0.0375	-0.0056
			(0.0569)		(0.0111)***	(0.0123)
		Num comments	-0.391	-0.1153	-0.0375	-0.0056
			(0.1963)*	(0.0553)*	(0.0111)***	(0.0123)
		Post word count	-0.5337	0.0934	-0.0375	-0.0056
			(0.4621)	(0.0824)	(0.0111)***	(0.0123)
		Max comment depth	-0.1592	0.0861	-0.0375	-0.0056
			(0.1234)	(0.0839)	(0.0111)***	(0.0123)
		Proactive	-0.2465	0.2475	-0.0375	-0.0056
			(0.2097)	(0.2032)	(0.0111)***	(0.0123)
		Includes URL	-0.0532	0.0317	-0.0375	-0.0056
			(0.0712)	(0.111)	(0.0111)***	(0.0123)
Pre-2021	False		0.0001		-0.0371	-0.0027
			(0.0146)		(0.0111)***	(0.0121)
		Num comments	0.0031	0.0011	-0.0371	-0.0027
			(0.0194)	(0.0081)	(0.0111)***	(0.0121)
		Post word count	0.0226	-0.0052	-0.0371	-0.0027
			(0.0229)	(0.0046)	(0.0111)***	(0.0121)
		Max comment depth	0.0124	-0.0088	-0.0371	-0.0027
			(0.0223)	(0.0146)	(0.0111)***	(0.0121)
		Proactive	0.0334	-0.0427	-0.0371	-0.0027
			(0.0387)	(0.0392)	(0.0111)***	(0.0121)
		Includes URL	-0.0094	0.0258	-0.0371	-0.0027
			(0.0188)	(0.0212)	(0.0111)***	(0.0121)
	True		0.0069		-0.037	-0.0027
			(0.013)		(0.0111)***	(0.0121)
		Num comments	0.0058	-0.0004	-0.037	-0.0027
			(0.0192)	(0.0082)	(0.0111)***	(0.0121)
		Post word count	0.0298	-0.0053	-0.037	-0.0027
			(0.0221)	(0.0047)	(0.0111)***	(0.0121)
		Max comment depth	0.01	-0.0023	-0.037	-0.0027
			(0.0225)	(0.0139)	(0.0111)***	(0.0121)
		Proactive	0.0325	-0.033	-0.037	-0.0027
			(0.0392)	(0.0388)	(0.0111)***	(0.0121)
		Includes URL	0.0024	0.0122	-0.037	-0.0027
			(0.0151)	(0.0172)	(0.0111)***	(0.0121)

Table 4.4: Panel regression results for the bullish post model where the dependent variable is the asset price change twelve weeks after the post. Standard errors are in parentheses, with * indicating levels of statistical significance.

Time	**Meme Stocks**	**Other Variable Name**	**Sentiment**	**Other Variable**	**CAR**	$\mathbf{r_{[t,t-1]}}$
All posts	Included		0.0158		-0.0512	-0.011
			(0.0312)		(0.0148)***	(0.0162)
		Num comments	0.0067	0.0042	-0.0512	-0.011
			(0.0265)	(0.0076)	(0.0148)***	(0.0162)
		Post word count	0.0672	-0.0135	-0.0512	-0.011
			(0.0191)***	(0.0038)***	(0.0148)***	(0.0162)
		Max comment depth	-0.0377	0.0398	-0.0512	-0.011
			(0.0181)*	(0.0275)	(0.0148)***	(0.0162)
		Proactive	0.0033	0.0183	-0.0512	-0.011
			(0.0445)	(0.0218)	(0.0148)***	(0.0162)
		Includes URL	0.0174	-0.0043	-0.0512	-0.011
			(0.0339)	(0.0155)	(0.0148)***	(0.0162)
	Excluded		-0.0224		-0.0513	-0.0109
			(0.0101)*		(0.0148)***	(0.0162)
		Num comments	-0.0319	0.0038	-0.0513	-0.0109
			(0.0189)	(0.0074)	(0.0148)***	(0.0162)
		Post word count	0.0497	-0.0156	-0.0513	-0.0109
			(0.0278)	(0.0058)**	(0.0148)***	(0.0162)
		Max comment depth	-0.0383	0.0117	-0.0513	-0.0109
			(0.0146)**	(0.0085)	(0.0148)***	(0.0162)
		Proactive	-0.09	0.0793	-0.0513	-0.0109
			(0.0288)**	(0.0268)**	(0.0148)***	(0.0162)
		Includes URL	-0.025	0.0071	-0.0513	-0.0109
			(0.0115)*	(0.0139)	(0.0148)***	(0.0162)
Post-2021	Included		0.0241		-0.0512	-0.0109
			(0.0481)		(0.0148)***	(0.0162)

Continued on next page

		Num comments	0.0105	0.0092	-0.0511	-0.0109
			(0.0297)	(0.0223)	(0.0148)***	(0.0162)
		Post word count	0.0599	-0.0133	-0.0512	-0.0109
			(0.0129)***	(0.0113)	(0.0148)***	(0.0162)
		Max comment depth	-0.0925	0.0927	-0.0511	-0.0109
			(0.0383)*	(0.0511)	(0.0148)***	(0.0162)
		Proactive	0.0131	0.0257	-0.0512	-0.0109
			(0.0478)	(0.0286)	(0.0148)***	(0.0162)
		Includes URL	0.0354	-0.0284	-0.0511	-0.0109
			(0.0624)	(0.0492)	(0.0148)***	(0.0162)
	Excluded		-0.0795		-0.0515	-0.0104
			(0.017)***		(0.0148)***	(0.0161)
		Num comments	-0.0852	0.0026	-0.0515	-0.0104
			(0.0314)**	(0.0127)	(0.0148)***	(0.0161)
		Post word count	-0.0112	-0.0147	-0.0515	-0.0104
			(0.0539)	(0.0111)	(0.0148)***	(0.0161)
		Max comment depth	-0.1014	0.0175	-0.0515	-0.0104
			(0.0289)***	(0.0189)	(0.0148)***	(0.0161)
		Proactive	-0.2773	0.2298	-0.0515	-0.0104
			(0.0591)***	(0.0607)***	(0.0148)***	(0.0161)
		Includes URL	-0.1023	0.0509	-0.0515	-0.0104
			(0.0218)***	(0.0291)	(0.0148)***	(0.0161)
Pre-2021	Included		0.0102		-0.0522	-0.0107
			(0.02)		(0.0148)***	(0.0162)
		Num comments	-0.0057	0.0061	-0.0522	-0.0107
			(0.0206)	(0.0096)	(0.0148)***	(0.0162)
		Post word count	0.1025	-0.0202	-0.0522	-0.0107
			(0.0447)*	(0.0072)**	(0.0148)***	(0.0162)
		Max comment depth	-0.011	0.0151	-0.0522	-0.0107
			(0.0158)	(0.0132)	(0.0148)***	(0.0162)
		Proactive	-0.0212	0.037	-0.0522	-0.0107
			(0.0332)	(0.0248)	(0.0148)***	(0.0162)
		Includes URL	0.009	0.0034	-0.0522	-0.0107
			(0.0194)	(0.0153)	(0.0148)***	(0.0162)
	Excluded		-0.0081		-0.0519	-0.0111
			(0.0116)		(0.0148)***	(0.0162)
		Num comments	-0.009	0.0004	-0.0519	-0.0111
			(0.0209)	(0.0083)	(0.0148)***	(0.0162)
		Post word count	0.0629	-0.0154	-0.0519	-0.0111
			(0.0299)*	(0.0061)*	(0.0148)***	(0.0162)
		Max comment depth	-0.0163	0.0058	-0.0519	-0.0111
			(0.0156)	(0.0096)	(0.0148)***	(0.0162)
		Proactive	-0.0465	0.0452	-0.0519	-0.0111
			(0.0281)	(0.0264)	(0.0148)***	(0.0162)
		Includes URL	-0.0086	0.0014	-0.0519	-0.0111
			(0.0125)	(0.0155)	(0.0148)***	(0.0162)

Table 4.5: Panel regression results for the bearish post model where the dependent variable is the asset price change twelve weeks after the post. Standard errors are in parentheses, with * indicating levels of statistical significance.

Time	**Meme Stocks**	**Other Variable Name**	**Sentiment**	**Other Variable**	**CAR**	$\mathbf{r_{[t,t-1]}}$
All posts	Included		-0.0098		-0.0513	-0.0084
			(0.0201)		(0.0148)***	(0.0159)
		Num comments	-0.018	-0.003	-0.0513	-0.0084
			(0.0226)	(0.0111)	(0.0148)***	(0.0159)
		Post word count	0.0488	-0.0137	-0.0513	-0.0084
			(0.0298)	(0.0076)	(0.0148)***	(0.0159)
		Max comment depth	-0.0219	0.0086	-0.0513	-0.0084
			(0.0279)	(0.0139)	(0.0148)***	(0.0159)
		Proactive	0.0165	-0.0351	-0.0513	-0.0084
			(0.0344)	(0.0318)	(0.0148)***	(0.0159)
		Includes URL	-0.0309	0.058	-0.0513	-0.0084
			(0.0234)	(0.0234)*	(0.0148)***	(0.0159)
	Excluded		0.0012		-0.0516	-0.0087
			(0.0195)		(0.0148)***	(0.0159)
		Num comments	-0.0343	-0.0129	-0.0516	-0.0087
			(0.0217)	(0.0093)	(0.0148)***	(0.0159)
		Post word count	0.0414	-0.0093	-0.0516	-0.0087

Continued on next page

			(0.0338)	(0.0073)	(0.0148)***	(0.0159)
		Max comment depth	-0.0172	0.0132	-0.0516	-0.0087
			(0.0276)	(0.0141)	(0.0148)***	(0.0159)
		Proactive	0.0354	-0.0439	-0.0516	-0.0087
			(0.0383)	(0.0363)	(0.0148)***	(0.0159)
		Includes URL	-0.0151	0.0438	-0.0516	-0.0087
			(0.0222)	(0.0227)	(0.0148)***	(0.0159)
Post-2021	Included		-0.093		-0.0518	-0.0106
			(0.0445)*		(0.0148)***	(0.0161)
		Num comments	0.1031	0.0793	-0.0518	-0.0105
			(0.1535)	(0.0858)	(0.0148)***	(0.0161)
		Post word count	0.141	-0.0622	-0.0518	-0.0106
			(0.0991)	(0.0416)	(0.0148)***	(0.0161)
		Max comment depth	-0.1093	0.0104	-0.0518	-0.0106
			(0.2298)	(0.1442)	(0.0148)***	(0.0161)
		Proactive	-0.0423	-0.1243	-0.0518	-0.0106
			(0.0322)	(0.0988)	(0.0148)***	(0.0161)
	Excluded		-0.0005		-0.0521	-0.0108
			(0.0729)		(0.0148)***	(0.0162)
		Num comments	-0.2574	-0.0841	-0.0521	-0.0108
			(0.1374)	(0.0542)	(0.0148)***	(0.0162)
		Post word count	0.1286	-0.0244	-0.0521	-0.0108
			(0.2413)	(0.0478)	(0.0148)***	(0.0162)
		Max comment depth	0.0101	-0.0076	-0.0521	-0.0108
			(0.1246)	(0.1226)	(0.0148)***	(0.0162)
		Proactive	0.0893	-0.107	-0.0521	-0.0108
			(0.0894)	(0.063)	(0.0148)***	(0.0162)
		Includes URL	-0.0493	0.1081	-0.0521	-0.0108
			(0.0492)	(0.1519)	(0.0148)***	(0.0162)
Pre-2021	Included		-0.0021		-0.0517	-0.0084
			(0.0201)		(0.0148)***	(0.0159)
		Num comments	-0.0282	-0.0095	-0.0517	-0.0084
			(0.0215)	(0.0092)	(0.0148)***	(0.0159)
		Post word count	0.0313	-0.0078	-0.0517	-0.0084
			(0.0348)	(0.0074)	(0.0148)***	(0.0159)
		Max comment depth	-0.0187	0.0119	-0.0517	-0.0085
			(0.0274)	(0.014)	(0.0148)***	(0.0159)
		Proactive	0.0309	-0.0424	-0.0517	-0.0084
			(0.0391)	(0.0377)	(0.0148)***	(0.0159)
		Includes URL	-0.0242	0.0597	-0.0517	-0.0085
			(0.0239)	(0.0254)*	(0.0148)***	(0.0159)
	Excluded		0.0013		-0.0516	-0.0085
			(0.0202)		(0.0148)***	(0.0159)
		Num comments	-0.026	-0.01	-0.0516	-0.0085
			(0.0216)	(0.0094)	(0.0148)***	(0.0159)
		Post word count	0.0407	-0.0092	-0.0516	-0.0085
			(0.0335)	(0.0074)	(0.0148)***	(0.0159)
		Max comment depth	-0.018	0.0138	-0.0516	-0.0085
			(0.0284)	(0.0139)	(0.0148)***	(0.0159)
		Proactive	0.0338	-0.0419	-0.0516	-0.0085
			(0.0394)	(0.0375)	(0.0148)***	(0.0159)
		Includes URL	-0.0138	0.0409	-0.0516	-0.0086
			(0.0231)	(0.0225)	(0.0148)***	(0.0159)

4.2.1 Sentiment

Note that the sign for the sentiment variable is flipped for the bearish models, so it is always the case that a positive coefficient for the sentiment variable implies that prices are increasing in the same direction the post creator predicted, while negative coefficients imply that prices moved in the opposite direction to the prediction. The signs for other variables is not reversed, so positive coefficients imply increasing asset prices in all model specifications, and vice-versa for negative coefficients.

The coefficient for sentiment was statistically significant in 19 percent of our models (717 configurations). There are distinct patterns to the model configurations where this occurs. Firstly, the parameter is rarely statistically significant in the bearish models. The only time this occurs in the displayed results is in Table 4.5, for the post-2021 model with meme stocks included and no other independent variables (effect size -0.93, $p < 0.05$).

Secondly, when sentiment is statistically significant, the coefficient is negative 76 percent of the time. In the bullish four week models shown in Table 4.2, the sentiment parameter is never positive and statistically significant. In the 543 configurations where sentiment is statistically significant and the coefficient is negative, 63 percent of these models include only the posts published in 2021. This matches a finding from [28], that asset prices tend to move in the opposite direction of DD predictions made in 2021. Figure 4.2.1 shows how the sentiment parameter varies when altering the time window of our model. Interestingly, as Figure 4.2.1 demonstrates, this finding only holds in the models where meme stocks are excluded. If meme stocks are included, returns are negative in the short-run but positive (albeit with high uncertainty) at time horizons longer than 6 weeks. This likely reflects the large number of meme stock DD posts in 2021, many of which performed well as the price of meme stocks skyrocketed.

Finally, in the 174 configurations where sentiment is statistically significant and the coefficient is positive, 172 of those configurations contain an additional independent variable. Examples of this phenomenon are apparent in Table 4.4 in some of the models where post word count is included as an additional feature.

4.2.2 Post word count

Post word count is statistically significant in around 21 percent of models (130 configurations) where it was included, but when it is statistically significant, the coefficient is

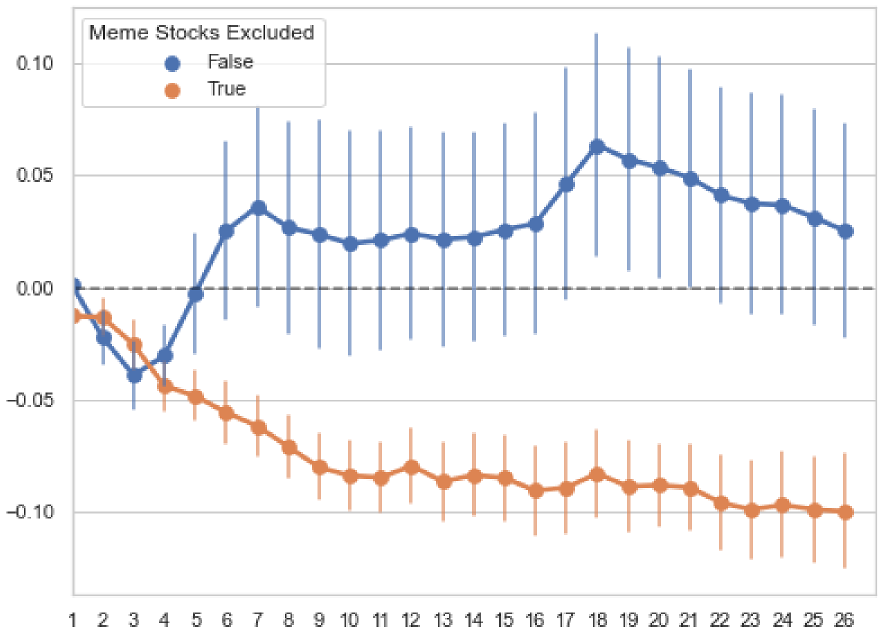


Figure 4.2.1: The coefficient for the sentiment variable for post-2021 bullish models with no additional features at different time horizons (1-26 weeks). Split by whether meme stocks (GME, AMC, BB, and NOK) are included. Bars indicate 95 percent confidence intervals.

negative 97 percent of the time. The vast majority of these configurations where word count is statistically significant are bullish models (116 configurations). For example, in Table 4.4, in the model including all bullish posts, the word count variable has an effect size of -0.0135 ($p < 0.001$). Notice, however, that in the same model, the sentiment parameter is statistically significant with an effect size of 0.0672 ($p < 0.001$), roughly five times the magnitude and the opposite sign. Since the number of comments variable was log-transformed, it turns out that the average post has a log-transformed word count of approximately five, meaning for the average post, the effects cancel each other out. Shorter DD posts therefore appear to be weakly predictive of future returns, although this effect is almost absent from the post-2021 models; only 11 of the configurations have a statistically significant post word count feature under this filter.

4.2.3 Number of comments and maximum comment depth

Our results indicate that the number of comments and the maximum comment depth are not associated with future price changes. Across all our configurations, number of comments and maximum comment depth were respectively statistically significant in 0.8

and 2.2 percent of the models they were included.

4.2.4 Proactive posts

The proactive variable is statistically significant in 27 percent of models. In bullish models where the variable is a statistically significant feature, the coefficient is always positive, and conversely in bearish models where the proactive variable is statistically significant, the coefficient is always negative. While this is promising, the actual effect size of the proactive variable is typically very close in magnitude to the sentiment variable with the opposite sign. This is demonstrated clearly in Figure 4.2.2. For example, in the four-week model with all bullish posts, sentiment has an effect size of -0.0404 ($p < 0.001$), while proactive has an effect size of 0.0443 ($p < 0.001$). These effects, therefore, are almost entirely cancelling each other out, which suggests that proactive posts have no association with future price changes, while reactive posts may have a slight negative association with future price changes.

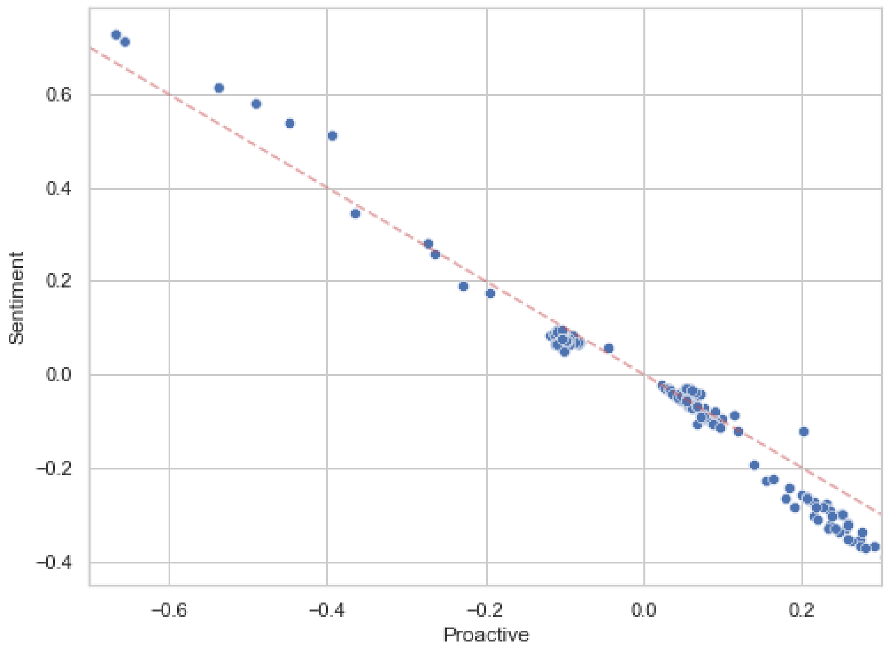


Figure 4.2.2: Scatterplot showing the parameter values of proactive and sentiment in all models where proactive is statistically significant. The red dashed line shows $y = -x$

4.2.5 Posts containing URLs

The URL feature is statistically significant in 18 percent of models, and always positive when significant. This feature tends to only be significant in the bearish models, for example, in the pre-2021 bearish model including meme stocks, the URL coefficient is 0.0597 ($p < 0.05$). Unlike the proactive post variable the effect is not simply washing out the sentiment effect, as the sentiment parameter for this model is negative with effect size of -0.0242 (not statistically significant). Since this is a bearish model, the results imply that bearish posts with URLs are more likely to precede an asset price increase. The URL parameter is also more likely to show up as positive for models looking at time horizons greater than 8 weeks, and especially at time horizons greater than 15 weeks.

CHAPTER 5

Discussion and Conclusion

5.1 Discussion

Our analysis suggests that WSB activity is not predictive of future stock returns, at least outside the context of 'meme' stocks where WSB activity has had a direct effect on market prices.

WSB overall sentiment appears to be a lagging indicator of future returns, not a leading one, i.e. WSB members appear to talk more about a stock when its price has increased rapidly. Our Granger causality tests showed that for the majority of stocks, WSB sentiment was not useful for forecasting future returns. Unsurprisingly, this is not true for stocks like GME or AMC, both of which underwent sharp price increases in response to retail trader activity emerging directly from the forum. This finding is, however, difficult to use in practise, since it would require a method of identifying when a stock starts entering a 'meme' period in real-time, which is a potential area for future research (see

[3] for some initial work on this problem). The only surprising exception we find is that WSB sentiment does seem to be useful for predicting short-term changes in the return of SNAP, though the reasons for this are unclear.

The results of the sentiment/return analysis largely cohere with [27], but diverge from [25]. The latter paper was one of the earliest analyses of this problem, and suffered from a few limitations – notably their calculated sentiment scores only involved simple word counts of the terms 'buy' and 'sell'. While our sentiment detection method is far from perfect accuracy, it does not rely on keywords, and therefore cannot be biased by the specific sentiment word list chosen by the authors.

When we restrict our analysis to the due diligence (DD) posts, the results suggests that the DD posts have either no association with price changes, or a negative association for posts after 2021 when meme stocks are excluded. This might suggest that a trader could make money in some cases by 'inverting' WSB advice, for instance by shorting stocks that the forum is bullish on. However, such a strategy would risk catastrophic losses, since the 2021 finding does not hold when meme stocks (GME, AMC, BB, and NOK) are included in the model.

One surprising finding is that longer bullish posts are more likely to precede a price decrease than shorter bullish posts, although this effect is not apparent for posts during 2021. One possible explanation is that longer posts might be working hard to justify investment decisions that are counter-intuitive – perhaps trying to make a particular stock reach meme status, or promoting a company with poor looking financials. On the other side, it is possible that when DD authors become aware of genuinely valuable information about a company that is not widely known to the public, they are more discrete in the details they share.

There is some evidence that bearish posts containing URLs are more likely to precede stock price increases over longer time horizons (8+ weeks). This effect could be explained by the fact that that bearish DD posts with URLs are sharing public information about why a company is struggling, which markets have priced in, but over-reacted to [41], and then the price recovers over sufficiently long time horizons.

Posts that were marked as proactive (i.e. where the CAR was close to zero at the time of submission) superficially appear to be more predictive than other kinds of post, but when considered against the parameter estimates for sentiment, the effect cancels out. This suggests that proactive posts are not associated with changes in price. Reactive posts, by contrast, appear to be weakly negatively associated with price changes. This result is somewhat surprising, given that we control for cumulative abnormal return, so this is not simply mean reversion. Digging into the data a little reveals a potential explanation is not simply WSB investors piling in on over-valued assets, but bullish bets made on stocks that were crashing but had not hit the floor. For example, one of the worst performing bets in the data was a bullish post about Waitr Holding's in August 2019. The stock was worth over 270 USD in March 2019, and was down to around 29 USD when the post was created. The stock continued to collapse in value, and was worth around 7 USD by the end of the year.

Our two graph-based features (number of comments and maximum comment depth) ended up being largely irrelevant in our analysis. Compared to the only other paper to look at DD posts directly [28], our results are aligned for the 2021 sample, but not for the pre-2021 sample. In the prior paper, the authors found that pre-2021, DD posts were positively associated with future returns. We suspect this is due to the two following factors: firstly, we likely have slightly different samples, which can matter a fair amount when we have particularly drastic changes in asset price over short periods of time (a fairly common occurence given WSB's penchant for high risk stocks). Secondly, there are differences in the specification of our models, for example which control variables are used, and our decision to separate bullish and bearish posts into different model configurations.

Despite these differences, we note that the coefficients observed by [28] for their net DD variables are only statistically significant at a 5 or 10 percent level (depending on their specification), and therefore could be a false positive. When considering the results of this paper and the results from [28] together, we would argue that if the pre-2021 DD posts are predictive of future returns, the magnitude of that effect might be slightly positive, but small.

Our research has a few limitations. Sentiment detection on a forum like WSB is difficult, and our model was not amazingly accurate, although it is a more sophisticated approach than used in previous papers. A more appropriate BERT model could certainly be developed by training a base model to perform masked word detection on the forum corpus before fine-tuning the model to perform sentiment detection on WSB posts.

The statistically significant negative results found consistently in the post-2021 bullish models when meme stocks are excluded reflects another potential limitation in the study regarding our choice of control variables. In the wake of the GME short-squeeze, many users were focused on identifying another stock where a similar phenomenon could be inculcated – AMC for example. One simple method to identify another potential candidate meme stock could be to look at the percentage of public stock that had been sold short (GME was notably one of the most shorted stocks on the market in late 2020). It is plausible that stocks with such high short-interest tend to be companies in particularly dire financial circumstances, and accordingly, we might expect them to generally under-perform. There are a litany of other potential control variables that are likely also relevant (e.g. changes in volume, put/call option ratios, news events) that could be added to the analysis to better discern the relationship between WSB advice and changes in underlying asset value.

Similarly, the use of the S&P 500 as an index for fitting the CAPM was a practical simple choice, but not necessarily optimal for providing the most accurate estimates of abnormal return. Given WSB users proclivity for liking technology companies, a Nasdaq 100 index like the Invesco QQQ might have been a more appropriate general choice. Other alternatives could have been to use an index that tracks stocks broadly similar to what WSB users are interested in (e.g. the ARK Innovation Fund), or using different market indices for different stocks. This could again help with improving the usefulness of the controls in the analysis, and making the effect estimates more accurate.

Another limitation is that the forum data only runs up to mid-June 2021, and it is quite possible that the nature of posts on the forum have changed significantly over the past two years. Updating our analysis with more recent posts would also help to expand our sample of DD posts, of which we only had a few thousand. Given that financial signals are highly noisy, it is quite possible that our DD models have insufficient statistical power to pick up meaningful effects.

There are additional features to WSB posts that have not been directly examined in this paper, but may be relevant. We considered the presence of URLs in posts, but our analysis does not attempt to identify the importance of the sources of URLs. Additionally, many posts contain images, which is a completely additional layer of content that we were not able to consider in this work. Additionally, the value of DD posts on Reddit may depend on authorship, or be revealed by the various social interactions of contributors on the platform. While we explored this area, engineering useful features our of the social network on Reddit proved challenging. Many comments and submissions are left by authors who have since deleted their accounts, and unlike Twitter, followers are not publicly visible, making it more difficult to identify, for example, users that have more prestige. Future analysis may have more success in engineering features in this space, as well as considering the other content-based features that we were not able to consider.

Perhaps the most interesting direction for future research would be to try to understand the dynamics of how a stock becomes a 'meme stock'. Our analysis suggests that in the majority of cases, WSB activity does not predict stock prices, but meme stocks are a clear exception, and the effect of the forum on the price of these assets is obvious. Being able to predict the emergence of meme stocks would have clear value in predicting future asset prices, but beyond that, it could yield genuinely insightful findings about social dynamics on online forums and their effects on markets.

5.2 Conclusion

The analysis of this paper suggests that while tracking WSB could be useful for identifying price-spikes that occur for meme stocks, it otherwise provides little to no valuable signal on future market prices, and retail traders following WSB advice are potentially making very risky decisions that could be very costly. Regulators should be aware that such forums could present a danger to retail investors who take the advice of the forum seriously.

Bibliography

[1] J. Ponciano, "Meme stock saga officially over? gamestop short interest plunged 70% amid $20 billion loss," Feb 2021.

[2] R. Armstrong, "R/wallstreetbets throws down challenge to regulators," Jan 2021.

[3] M. Costola, M. Iacopini, and C. R. Santagiustina, "On the "mementum" of meme stocks," *Economics Letters*, Vol. 207, p. 110021, 2021.

[4] C. Wang and B. Luo, "Predicting $ gme stock price movement using sentiment from reddit r/wallstreetbets," in *Proceedings of the Third Workshop on Financial Technology and Natural Language Processing*, pp. 22–30, 2021.

[5] X. Wan, J. Yang, S. Marinov, J.-P. Calliess, S. Zohren, and X. Dong, "Sentiment correlation in financial news networks and associated market movements," *Scientific Reports*, Vol. 11, no. 1, p. 3062, 2021.

[6] H. K. Sul, A. R. Dennis, and L. I. Yuan, "Trading on twitter: Using social media sentiment to predict stock returns," *Decision Sciences*, Vol. 48, no. 3, pp. 454–488, 2017.

[7] V. S. Pagolu, K. N. Reddy, G. Panda, and B. Majhi, "Sentiment analysis of twitter data for predicting stock market movements," in *2016 International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES)*, pp. 1345–1350, 2016.

[8] T. Rao, S. Srivastava, *et al.*, "Analyzing stock market movements using twitter sentiment analysis," 2012.

[9] S. D. Tan and O. Tas, "Social media sentiment in international stock returns and trading activity," *Journal of Behavioral Finance*, Vol. 22, no. 2, pp. 221–234, 2021.

[10] R. Gao, Z. Zhang, Z. Shi, D. Xu, W. Zhang, and D. Zhu, "A review of natural language processing for financial technology," in *International Symposium on Artificial Intelligence and Robotics 2021*, Vol. 11884, pp. 262–277, SPIE, 2021.

[11] E. F. Fama, "Efficient capital markets: A review of theory and empirical work," *The Journal of Finance*, Vol. 25, no. 2, pp. 383–417, 1970.

[12] H. Chen, P. De, Y. J. Hu, and B.-H. Hwang, "Wisdom of Crowds: The Value of Stock Opinions Transmitted Through Social Media," *The Review of Financial Studies*, Vol. 27, pp. 1367–1403, 02 2014.

[13] J. B. De Long, A. Shleifer, L. H. Summers, and R. J. Waldmann, "Noise trader risk in financial markets," *Journal of Political Economy*, Vol. 98, no. 4, pp. 703–738, 1990.

[14] F. BLACK, "Noise," *The Journal of Finance*, Vol. 41, no. 3, pp. 528–543, 1986.

[15] M. Farrell, T. C. Green, R. Jame, and S. Markov, "The democratization of investment research and the informativeness of retail investor trading," *Journal of Financial Economics*, Vol. 145, no. 2, Part B, pp. 616–641, 2022.

[16] J. A. Cookson, J. Engelberg, and W. Mullins, "Echo chambers," *Available at SSRN 3603107*, 2022.

[17] W. Jia, G. Redigolo, S. Shu, and J. Zhao, "Can social media distort price discovery? evidence from merger rumors," *Journal of Accounting and Economics*, Vol. 70, no. 1, p. 101334, 2020.

[18] S.-H. Kim and D. Kim, "Investor sentiment from internet message postings and the predictability of stock returns," *Journal of Economic Behavior Organization*, Vol. 107, pp. 708–729, 2014. Empirical Behavioral Finance.

[19] J. Davidson, "Meet the bros behind /r/wallstreetbets, who lose hundreds of thousands of dollars in a day—and brag about it," Oct 2018.

[20] "Subreddit stats."

[21] P. K. Schou, E. Bucher, M. Waldkirch, and E. Grünwald, "We did start the fire:

r/wallstreetbets, 'flash movements' and the gamestop short-squeeze," *Academy of Management Proceedings*, Vol. 2022, no. 1, p. 14028, 2022.

[22] A. Anand and J. Pathak, "Wallstreetbets against wall street: The role of reddit in the gamestop short squeeze," 2021.

[23] C. Boylston, B. Palacios, P. Tassev, and A. Bruckman, "Wallstreetbets: positions or ban," *arXiv preprint arXiv:2101.12110*, 2021.

[24] J. Winkler and V. Semenova, "Reddit's self-organised bull runs: Social contagion and asset prices," INET Oxford Working Papers 2021-04, Institute for New Economic Thinking at the Oxford Martin School, University of Oxford, Jan. 2021.

[25] T. Buz and G. de Melo, "Should you take investment advice from wallstreetbets? a data-driven approach," *arXiv preprint arXiv:2105.02728*, 2021.

[26] D. W. Witts, E. Tortosa-Ausina, I. Arribas, *et al.*, "The irrational market: Considering the effect of the online community wall street bets on financial market variables," tech. rep., 2021.

[27] R. G. Chacon, T. G. Morillon, and R. Wang, "Will the reddit rebellion take you to the moon? evidence from wallstreetbets," *Financial Markets and Portfolio Management*, pp. 1–25, 2022.

[28] D. Bradley, J. Hanousek Jr, R. Jame, and Z. Xiao, "Place your bets? the market consequences of investment advice on reddit's wallstreetbets," *The Market Consequences of Investment Advice on Reddit's Wallstreetbets (March 15, 2021)*, 2021.

[29] J. Baumgartner, S. Zannettou, B. Keegan, M. Squire, and J. Blackburn, "The pushshift reddit dataset," in *Proceedings of the international AAAI conference on web and social media*, Vol. 14, pp. 830–839, 2020.

[30] C. Hutto and E. Gilbert, "Vader: A parsimonious rule-based model for sentiment analysis of social media text," *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 8, pp. 216–225, May 2014.

[31] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," *arXiv preprint arXiv:1706.03762*, 2017.

[32] D. Araci, "Finbert: Financial sentiment analysis with pre-trained language models," *arXiv preprint arXiv:1908.10063*, 2019.

[33] P. Malo, A. Sinha, P. Takala, P. Korhonen, and J. Wallenius, “Good debt or bad debt: Detecting semantic orientations in economic texts,” 2013.

[34] E. F. Fama and K. R. French, “The capital asset pricing model: Theory and evidence,” *Journal of Economic Perspectives*, Vol. 18, pp. 25–46, September 2004.

[35] C. W. J. Granger, “Investigating causal relations by econometric models and cross-spectral methods,” *Econometrica*, Vol. 37, no. 3, pp. 424–438, 1969.

[36] S. Vosoughi, D. Roy, and S. Aral, “The spread of true and false news online,” *Science (New York, NY)*, Vol. 359, no. 6380, pp. 1146–1151, 2018.

[37] N. Rosenfeld, A. Szanto, and D. C. Parkes, “A kernel of truth: Determining rumor veracity on twitter by diffusion pattern alone,” in *Proceedings of The Web Conference 2020*, pp. 1018–1028, 2020.

[38] F. Monti, F. Frasca, D. Eynard, D. Mannion, and M. M. Bronstein, “Fake news detection on social media using geometric deep learning,” *arXiv preprint arXiv:1902.06673*, 2019.

[39] E. O. Thorp, “The kelly criterion in blackjack sports betting, and the stock market,” in *Handbook of asset and liability management*, pp. 385–428, Elsevier, 2008.

[40] J. M. Poterba and L. H. Summers, “Mean reversion in stock prices: Evidence and implications,” *Journal of Financial Economics*, Vol. 22, no. 1, pp. 27–59, 1988.

[41] P. Veronesi, “Stock market overreaction to bad news in good times: A rational expectations equilibrium model,” *The Review of Financial Studies*, Vol. 12, no. 5, pp. 975–1007, 1999.