

WallStreetBets: An Analysis of Investment Advice Democratization

Tolga Buz
Hasso Plattner Institute, Potsdam, Germany
tolga.buz@hpi.de

Gerard de Melo
Hasso Plattner Institute, Potsdam, Germany
gdm@demelo.org

Abstract

Reddit's WallStreetBets (WSB) community has come to prominence due to its role in the hype around GameStop and other meme stocks. Yet very little is known about the reliability of the investment advice disseminated on WSB. We investigate whether an anonymous, investment-focused community such as WSB can be a valuable source for investment advice and thus may constitute a way of democratizing access to financial knowledge. Our analysis reviews data spanning 28 months to assess how successful an investor relying on WSB recommendations could have been. We find that the portfolio of most discussed stocks has grown significantly, outperforming the S&P 500 over the reviewed time frame. We further find that following buy signals at the time they are posted on WSB leads to positive outcomes over the long run, and that the GameStop hype merely amplified previously existing characteristics.

Keywords: open finance, social media, collaborative investing, text mining, collective intelligence

1. Introduction

Members of Reddit's *r*/WallStreetBets (WSB) community often make a point to emphasize that their posts should not be deemed as constituting financial advice. Yet, the forum appears to have profoundly affected a number of specific stock prices, most prominently in the January 2021 GameStop short squeeze. Some of these have since been known as *meme stocks*, reflecting the social contagion arising from trending posts in the forum. The community's size and reach likely contributed to its important role during the GameStop hype: *r*/WallStreetBets is currently the largest trading-focused online community and has drawn substantial attention to social communities on platforms such as Reddit. Unlike traditional online services, the WSB community's language is notably profane and users are often ridiculed when posting their

losses and analyses, yet they appear united by a common goal of yielding high profits on the stock market. In contrast to financial news outlets and online publishers (e.g., The Motley Fool) as well as reports published by investment banks or prominent individuals on TV or social media, a topic-focused community such as WSB offers a more democratic environment for obtaining investment signals: WSB's posts are freely accessible, contributed by a large number of users, and filtered for quality to a certain degree using WSB's guidelines and Reddit's voting mechanism. While WSB is widely known for its highly speculative and unorthodox forms of sharing and debating investment strategies, little is known about the reliability of the disseminated recommendations. WSB users are well aware of and often mention the constant risk of "pump and dump" schemes, of simply being too late on a trend, or of manipulation through malicious third parties, which can lead to financial losses and "bag holding".

Can the social dynamics in a community of anonymous traders generate investment advice that outperforms the broader market? In an attempt to shed light on this question, we present an analysis based on WSB posts spanning a time period that includes the two years prior to the GameStop hype of 2021 as well as the year following the COVID-19-related stock market crash in spring 2020. We present new insights from a data-driven perspective on the WSB community's proficiency in order to evaluate how well investors following the submissions and buy recommendations in the forum would have performed financially in this period. The stock market is a complex system affected by a multitude of factors. We thus do not aim to answer how and to what extent WSB might have the power to move financial markets, as may have occurred with some of the meme stocks. Instead, we investigate to what extent specific patterns in the posts on *r*/WallStreetBets may have anticipated corresponding stock price movements that occurred later, by acquiring a large WSB dataset and enriching it with financial data. We establish techniques to identify the most popular

stocks within thousands of text snippets, identify the portfolio of most discussed stocks, and assess how reliable the community's buy and sell signals would have been if taken as investment advice. Our analysis shows that not every single investment decision shared on WallStreetBets leads to financial success, yet the signals achieve average investment returns that can significantly outperform the S&P 500 index, which we here consider as a proxy for the broader market. We further develop means to distinguish different sorts of posts and identify proactive buy signals with promising prospects. Additionally, we compare analysis results based on data from the *pre-hype* and *post-hype* time frames, as the rapid rise to prominence of the WSB community in January 2021 also entailed an explosive growth in user numbers and generated content.

Our work contributes to understanding the changing role of WSB and similar online communities in our society. Despite being designed as a medium for leisurely entertainment and informal exchanges, we show that such communities can have an impact on the personal and professional lives of their large user base by offering a valuable source of information and democratic access to investment advice that otherwise requires trusting single sources of truth, and, in many cases, paying large amounts of money for access.

2. Background

2.1. Reddit, WSB, and the Short Squeeze

Reddit was established in 2005 and is a popular social media platform that provides a place for communities (known as *subreddits*) on specific topics, such as humorous memes, politics, or computer games. The platform has since grown to host over 100,000 subreddits with more than 52 million daily active participants, accumulating more than 50 billion monthly views (according to redditinc.com). Members can post submissions, comment on them, and use up-/down-vote buttons to rate submissions and comments.

WallStreetBets (WSB) is a subreddit created on January 31, 2012 and accessible via www.reddit.com/r/wallstreetbets. WSB's number of subscribers increased rapidly from 1.7 million to approximately 8.5 million within January 2021 (source: subredditstats.com), due to the media attention received in the context of the GameStop hype. At the time of writing, WSB has more than 11 million subscribers. The website subredditstats.com lists WSB at rank 49 out of all subreddits in terms of subscriber counts, rank 13 with respect to the number of comments per day (in March/April 2021 even second rank),

and rank 32 regarding the number of posts per day, indicating a highly active community. WSB has the highest subscriber count of all finance-related Reddit communities. It describes itself as "*a community for making money and being amused while doing it. Or, realistically, a place to come and upvote memes when your portfolio is down.*" WSB is notorious for its unique slang and the pervasive use of offensive terms, e.g., WSB subscribers are officially referred to as "degenerates" and often call each other "apes" or "retards". Notable differences have been found between WSB and other finance-related communities on Reddit (Agrawal et al., 2022).

WSB saw an unprecedented rise in popularity and news coverage when the community started debating a particular set of stocks, including GameStop, that were in part deemed undervalued while simultaneously exhibiting a high short interest, i.e., the ratio of shares being sold short (or *shorted*) by financial institutions. Short-selling a stock refers to the practice of borrowing shares of a stock in order to sell them immediately with the goal of buying them back later at a lower price. This practice is a speculative move often employed by professional investors on stocks that are considered overvalued and expected to lose value in the near future. However, if the stock price instead increases, a short position can lead to considerable losses.

Discussing potentially undervalued investment opportunities is common in investment-focused communities, but GameStop's high short interest led to WSB's situation being portrayed as an ideological David-and-Goliath battle of small retail investors rallying against hedge funds: by buying and holding the stock, its price could be driven up, thus forcing the financial institutions to close their short positions at a significant loss, which in turn drove the stock price up even further – a phenomenon known as a *short squeeze*. This made WallStreetBets a place where the broader community of Reddit users could come together to unite in a movement, driven by the prospect of large financial gains through risky investments, with the added appeal of supposedly advancing the greater cause of punishing the financial institutions, particularly hedge funds. The latter are accused by the WSB community of having ruined many people's lives during the financial mortgage crisis of 2008, among other events.

2.2. Related Research

Social media can be a valuable source of data for assessing the public sentiment on topics such as politics, companies, brands, and products, and is frequently studied especially in behavioural finance and decision

sciences. Research suggests that there are important ties between social media sentiment and macroeconomic factors such as consumer confidence (Daas & Puts, 2014) and that the general social media sentiment may help understanding a firm's stock performance (Yu et al., 2013). Extensive research has attempted to show how particular cues from social media can enable predicting stock price changes (Duz Tan & Tas, 2021; Nguyen et al., 2015; Sul et al., 2017). Another line of work focuses on the forms of interaction in modern online platforms. For instance, emotionally charged messages in social media have been shown to tend to spread more quickly (Stieglitz & Dang-Xuan, 2013). Messages disseminated on social media, in some circumstances driven by bots, have empowered social movements (Manikonda et al., 2018) and influenced politics at an international scale (Gorodnichenko et al., 2018; Howard et al., 2011). More general research has reviewed the effects of social media on different target groups, and how they access and utilize its information during their daily lives (Hajli, 2014; Kim et al., 2014). The theory of collective intelligence also supports the ability of online communities to achieve common goals through participation and collaboration (Leimeister, 2010; Preece & Shneiderman, 2009).

However, very little research is available on the topic of how social media can be harnessed for information acquisition and how it makes knowledge more accessible (Tajudeen et al., 2016). Fuelled by the 2021 GameStop hype on WallStreetBets, recent research has provided a number of insights on the dynamics and background of the matter. This includes studies on the social dynamics within the WSB community that led to the hype (Lyócsa et al., 2022; Schou et al., 2022) and on the idea of retail investors fighting against Wall Street (Chohan, 2021; H. Jones & Hietanen, 2022). Other studies considered the financial mechanisms driving the sudden price spike (Anand & Pathak, 2021; Wang & Luo, 2021) and the effect of retail traders on prices and volatility, along with their participation in transactions (Hasso et al., 2022). Further research assessed the implications of the events for market regulators and brokerages (C. M. Jones et al., 2021; Umar et al., 2021). One study conducted an analysis of selected posts from an anthropological perspective (Mendoza-Denton, 2021). Several studies investigated the social interaction on WSB and explained the nature of the community's conversations and language (Agrawal et al., 2022; Boylston et al., 2021). Most existing research focuses on socio-economic and general market effects and implications. Little is known about the financial skills of WSB traders and the merits of their investment advice (Bradley et al., 2021). There is also a lack of research

from an information science perspective studying data that span a longer time frame, a large number of different postings, and assessing longer-term effects and trends.

3. Dataset Compilation

3.1. Data Acquisition

The data considered in our study was collected programmatically by accessing the Pushshift Reddit archiving service Baumgartner et al. (2020). The latter allows requesting data exports in JSON format based on submissions or comments. In order to compile a representative dataset, submissions from January 1, 2019 to April 4, 2021 were retrieved in a first step. This covers a time span to analyze the activity on WSB starting two years before the January 2021 GameStop incident until a year after the COVID-19 pandemic affected stock markets in March/April 2020. We employ an upvote score threshold to reduce noise, as posts with an upvote-to-downvote ratio below 50% do not offer valuable content – these submissions were often deleted by the author or a moderator, do not meet quality or community standards, are duplicate posts created by bots, or are unfit in some other way. This ensures that the dataset is representative of what community members normally see, as the above kinds of posts do not appear in a WSB visitor's standard view.

The data provided by the Pushshift API include additional metadata, e.g., the time of posting, the author identifier, a potential category tag ("flair"), etc. While the service also offers the option of collecting comments provided in response to submissions, in this analysis we focus on submissions only, because they represent the main topics that a user encounters when visiting and browsing WallStreetBets. Comments are typically short replies to the topic presented in the submission, but are not as prominently displayed and very heterogeneous in terms of their information value. It should be noted that the Reddit data provided by the Pushshift service may deviate from the original data in minor ways. For some posts, the upvote scores provided by the API are lower than the true value. Additionally, in the beginning of February 2021, the service was offline for multiple days. We acknowledge the risk that due to similar outages in the past, Pushshift may not have been able to scrape Reddit consistently. However, other relevant metadata, e.g., creation date and number of comments, were always correct in our sample comparison. Thus, we conclude that apart from the score value, the data is sufficiently reliable to conduct the analysis.

3.2. Dataset Statistics

The overall dataset consists of 778,288 submissions, of which 238,001 include a body text (while those lacking a body text may have included an image or video instead). A significant portion of these submissions was posted in 2021 (494,885 in total, 124,704 with body text), due to the aforementioned dramatic expansion of the community's subscriber count and amount of posted content. This shows how much traction the subreddit gained in 2021, in large part due to the GameStop hype, and is in line with the growth of subscribers from 1.7M to 8.5M within January 2021. In general, the vocabulary sizes are fairly large, which in part stems from the fact that the corpus consists of user-generated text from social media and hence harbours substantial numbers of typos, emojis, symbols, and similar phenomena that can increase the vocabulary size. On WSB, a system of tags, known as link *flairs* (Explained in the guidelines accessible via <https://www.reddit.com/r/wallstreetbets/wiki/linkflair>), serve as labels to categorize posts as *Discussion*, *DD* (due diligence), *News* (concerning specific stocks or markets), *Gain* and *Loss* (recounting gains or losses that need to surpass a certain monetary value threshold in order to be approved), or *YOLO* (high-risk high-value investments), among others. In our dataset, 718,517 submissions are labelled with such a tag either by their author or by the moderators. A review of this feature shows the following distribution: 23.5% *Discussion*, 22.1% *Meme*, 12.6% *YOLO*, 8.6% *News*, 7.7% *Gain*, 7.6% *Shitpost*, 4.9% *Loss*, 4.5% *DD*, and 8% others. This suggests that WSB serves two functions: for discussing investment ideas and news on one hand, and entertainment on the other.

4. The WallStreetBets Portfolio

4.1. Methodology: Detecting Stock Tickers

Each stock can be identified by its ticker, which is usually an abbreviation related to the company, conventionally written in capital letters, and sometimes adorned with a prefixed "\$" (e.g., \$AAPL for Apple Inc.) in order to identify them as stock tickers – this is particularly helpful when it could otherwise be mistaken for a regular word or abbreviation (e.g., \$COST for Costco Wholesale Corp.). Unfortunately, authors frequently omit the preceding dollar sign, in many cases use lowercase letters for the tickers, or write entire sentences or posts in uppercase letters. Additionally, many common abbreviations of business- or sector-related terms can be mistaken for stock tickers, e.g., GM for General Motors Co. or General Manager. Sometimes, posts about a particular stock

may include mentions of additional companies, e.g., if the bank JPMorgan Chase & Co. (JPM) has commented on buying or selling a specific stock. This makes it challenging to detect stock tickers correctly without false positives or false negatives. We adopt the following technique: After tokenization and removing punctuation, we filter for short tokens with 2–5 uppercase alphabetic characters or tokens starting with "\$" followed by 1–5 alphabetic characters. Single-character stock tickers are considered only with a "\$" to avoid false positives, e.g., \$F is the stock ticker for Ford Motor Co., while "F" usually refers to a popular swear word. Tokens without an affixed dollar sign are counted if they are included in the list of 7,100 stocks in the Stock Screener available on NASDAQ's website or in a list of ETF tickers (extracted manually from the WSB community's discussions about ETFs), while not part of a list of exceptions (a list of short tokens that are often used as words or abbreviations and could be mistaken as stock symbols). Our evaluation confirms that the WSB community frequently includes a "\$" to specify stock tickers that could be taken as normal words or are single letters. While a high-precision method to detect stock tickers could be to focus on tickers mentioned with a preceding dollar sign only, this would have very low recall - e.g., only 13.4% of all \$GME (GameStop) and 11.6% of all \$MSFT (Microsoft) mentions had a preceding "\$". The strategy proposed above instead constitutes a method of detecting the vast majority of stock tickers, while minimizing the amount of false positives and false negatives.

4.2. Analysis

Invoking the technique described above, we identify the 100 most frequently mentioned stock and ETF tickers for different time windows: the dataset's full range, as well as each year separately (2019, 2020, 2021). Grouping the tickers contained in these lists by sector shows that the WSB community focuses primarily on stocks from the sectors *Consumer Cyclical* (non-essential consumer goods), *Technology* (manufacturers of tech hardware and software), *Healthcare* (including pharmaceuticals, cannabis companies), and *Communication Services* (including interactive media, entertainment, telecommunications), as reported in Table 1. The distribution of the sectors' market capitalization shows that WSB appears to focus on sectors that are close to consumers.

Particular attention is given to stocks of younger companies, which are often considered growth stocks. These tend to be riskier, but exhibit a higher growth potential (in contrast to low-volatility value stocks such

as The Coca-Cola Company): The three sectors that performed better than the S&P 500 index regarding their 3-year growth (*Technology*, *Consumer Cyclical*, and *Communication Services*) are all ranked within the community's top four sectors. Regarding the 1-year growth, WSB's focus on the same three sectors and *Industrials* and *Financial services* outperformed the S&P 500 as well, although the community seems to have missed out on the strong growth in *Basic Materials* and *Energy*. The S&P 500 is considered a good reference for the broader (U.S.) market. However, there has also been a focus on the healthcare industry (mainly drug manufacturers, especially companies selling cannabis or COVID-19 vaccines), which performed worse than the broader market over the 3-year and 1-year time windows.

Table 1. Distribution of top 100 WSB-mentioned tickers per sector per year with market capitalization, 1-year, and 3-year change (as of March 2021)

Sector	Count per year			Financial data		
	'21	'20	'19	Cap.	1Y-Chg.	3Y-Chg.
Cons. Cyc.	22	24	20	\$8.9T	87.8%	58.5%
Tech	17	22	24	\$13.3T	75.7%	86.6%
Health	14	11	13	\$7.4T	37.4%	35.1%
Com. Srv.	9	10	16	\$6.2T	68.7%	54.4%
Industrials	11	11	6	\$5.5T	79.7%	26.3%
Fin. Srv.	10	5	6	\$8.1T	79.3%	18.0%
Cons. Def.	1	5	6	\$4.3T	19.0%	23.4%
Materials	5	2	0	\$2.6T	89.5%	31.7%
Energy	1	2	2	\$2.6T	105.7%	-25.7%
Utilities	0	2	2	\$1.5T	13.3%	23.0%
Real Estate	4	0	0	\$1.5T	33.1%	23.9%
S&P 500					63.2%	42.2%

We can conclude that WSB's focus on stocks from consumer- and technology-related sectors correspond to those that have generally seen beneficial outcomes in terms of the longer-term sector growth compared to the broader market performance. This conclusion is supported when taking a closer look at the most popular stocks and their products. We define the "WSB portfolio" based on the 21 tickers that have consistently been part of the 100 most frequently discussed stocks and ETFs on WSB for each of the three years.

The list is given in Table 2. We can group these stocks based on the respective industry: Entertainment & Internet (AMC, FB), Auto Manufacturers & Retail (GM, NIO, TSLA, AMZN, BABA, GME), Consumer Technology & Semiconductors (AAPL, AMD, INTC, NVDA), Technology & Software Infrastructure (BB, NOK, MSFT), Cannabis Companies (ACB, TLRY), and Industrials (GE, BA), ETFs (SPY, QQQ).

The detailed results in Table 2 reveal that 10 out of 18 tickers (NIO and TLRY went public at a later

time) performed better than the S&P 500 ETF (\$SPY) with respect to the 3-year change, and 14 out of 20 outperformed the S&P 500 ETF with respect to the 1-year change – both as of March 20, 2021. 11 of the 18 stocks (excluding the ETFs) are part of the S&P 500 index. While \$SPY achieved 51.93% growth over three years and 72.96% over the final year, a hypothetical, equally-weighted portfolio of these most frequently and consistently discussed stocks from WSB grew by 198.60% (mean) / 72.53% (median) over the final three years and by 483.20% (mean) / 103.74% (median) over the final year. Thus, the WSB portfolio grew faster than the broader market on average – particularly following the market crash caused by the COVID-19 pandemic. In conclusion, this means that during the longer term of two to three years as well as throughout the final year, investing in the most popular stocks on WSB most likely would have been a good investment decision (compared to overall market growth).¹

Table 2. Consistently and frequently discussed stock tickers with 1-year and 3-year price development (as of March 20, 2021)

Ticker	Company name	1Y-Chg. (%)	3Y-Chg. (%)
AAPL	Apple	110.99	183.88
ACB	Aurora Cannabis	15.75	-89.29
AMC	AMC Entertainment	336.68	20.10
AMD	AMD	99.60	611.61
AMZN	Amazon.com	66.57	93.82
BA	Boeing Co.	169.26	-20.88
BABA	Alibaba Group	32.26	20.53
BB	BlackBerry	223.96	-16.28
FB	Facebook	93.76	72.53
GE	General Electric Co.	103.74	3.82
GM	General Motors Co.	229.77	76.24
GME	GameStop Corp.	5226.33	1446.41
INTC	Intel Corp.	42.74	33.10
MSFT	Microsoft Corp.	69.41	157.43
NIO	NIO	1706.25	N/A
NOK	Nokia Corp	51.13	-25.67
NVDA	NVIDIA Corp	150.00	107.47
QQQ	Invesco QQQ Trust	84.49	90.68
SPY	SPDR S&P 500 ETF	72.96	51.93
TLRY	Tilray, Inc.	594.52	N/A
TSLA	Tesla, Inc.	665.88	954.37
Mean		483.20	198.60
Median		103.74	72.53

5. Investment Advice Reliability

Having assessed the overall portfolio, we now consider the reliability of particular investment advice at a more fine-granular level, investigating whether

¹The investment's timing is critical: the 1-year change coincides with the low-point of stock prices right after the COVID-19-related market crash in March 2020, which is why many 1-year performances beat the 3-year values.

specific types or intensity levels of WSB activity *at particular points in time* indicate a good timing for an investment decision. To this end, we evaluate how the value of the investment developed over specific short-term time frames, after a hypothetical investor decides to follow a WSB buy signal for a specific stock on a specific day. We assess price developments at specific points in time relative to the date in question, e.g., after one day, after three days, after a week, etc.

5.1. Methodology

In order to measure stock-related activity on WallStreetBets, multiple features are derived from the underlying data. As a baseline signal, we count mentions of a stock ticker or company name. For a more targeted assessment we count transaction-related words that are mentioned in the same submission as the respective ticker mentions, specifically “buy”, “hold”, “sell”, “call”, and “put” (and their variations), while excluding their negated forms, e.g., “not buy” and “don’t buy”. The collected word and ticker counts of each WSB post are aggregated per day in order to create a daily consensus for each stock ticker. This is then enriched with financial data taken from Yahoo! Finance with each stock’s daily closing price, high and low, and trading volume. We calculate a relative volatility as the difference between a day’s high and low values with regard to the closing price. As the stock markets are closed on weekends and public holidays, the volume and volatility on those days are set to zero and values for closing price and daily high and low are taken from the previous day. This helps in assessing time windows that fall on closed-market days – e.g., if someone bought a specific stock on a Friday at a price of \$30 and 30 days later on Sunday the closing price was at \$40, this is considered a positive price development. This results in a dataset consisting of WSB signals as well as financial data at a daily granularity level for each selected stock.

For a more detailed analysis, we computer a number of indicators. For each day, buy and sell mentions are compared – whichever count is higher defines whether the day is regarded as a *buy signal* or *sell signal*, which is treated as a piece of investment advice in this context. Additionally, for every day, we compute relative changes of the closing price per respective date compared to 1 day, 3 days, and 1 week before, as well as 1 day, 3 days, 1 week, 1 month, and 3 months later (time window x). Based on these price differences, we add various Boolean features to identify specific cases, e.g., price increased or decreased since/after x , Buy and Sell accuracy (a buy signal is considered accurate if the price increased after x). With these features, we assess

the success rate of WSB’s signals for each time window x : For an evaluation of signal accuracy, all signals per selected stock ticker are tracked with the respective price changes before and after in order to determine whether an investment made on a specific day would have gained or lost in value. A buy signal is considered successful within a specific time window if and only if the closing stock price increased within the same window – if it decreases or stays equal, the buy signal is considered a failure. For sell signals, the converse applies. We create such daily data summaries with WSB signals and financial data for each of the WSB portfolio stocks as well as for the overall S&P 500 index.

Note that we do not assume that all stock price movements are reflected in the WallStreetBets community’s discussions, let alone caused by the community – the stock market is a complex system affected by countless factors, of which one might be the WSB community. One may consider WSB activity that *precedes* significant changes in the stock market as indicative of WSB being a good source of information, even if the primary cause lies somewhere else and WSB only picked up on the development. However, the success of some pieces of information could also be due to external effects in the market rather than the community’s proficiency.

5.2. Short-Term Performance

On average, the WSB portfolio stocks have grown 3–6 times more than the S&P 500 – if an investor bought shares of the WSB portfolio on every day between January 2019 and December 2020, their investments would have grown by more than 16% on average after 3 months, while the S&P 500 would have grown slightly less than 5% (see Table 3). Indeed, the table shows that if an investor had chosen to buy on days of WSB activity regarding the respective portfolio tickers, the average 3-month growth would have reached 27.32% (buying if the ticker was mentioned), 32.03% (if additionally a buy signal was detected), or 28.53% (when buying despite a sell signal). Similar patterns can be observed for the shorter time frames of a day, a week, and a month. Overall, this reveals that on average the WSB portfolio outperformed the S&P 500 in short-term investment windows since 2019, if the timing of the investments followed WSB’s activity or signals. It should be noted that sell signals have been quite unsuccessful within the reviewed time windows – interpreting sell signals as buy signals would have generated a similarly beneficial outcome as relying on actual buy signals, with even higher average price increases for short term windows of one day and one week. This can be due to two

reasons: The market has been in a general upwards trend invalidating many sell signals at some point, and selling stocks at the right time is much more difficult than buying them, as prices tend to go down in much shorter time frames and more abruptly than they go up.

Table 3. Average price change per time window, mentions, daily volatility, and daily volume – for S&P 500 and WSB portfolio (Jan 2019 – Apr 2021)

	after	S&P	WSB Portfolio when bought on:			
		500	All	Mention	Buy Sig.	Sell Sig.
Price chg. (%)	1d	0.07	0.17	0.29	0.29	0.65
	1w	0.44	1.24	2.07	2.19	3.77
	1m	1.79	5.14	7.77	9.88	8.12
	3m	4.89	16.65	27.32	32.03	28.53
Avg. ment.		–	28.28	51.69	89.50	15.53
Avg. volat.		0.8%	2.9%	3.7%	4.2%	3.7%
Avg. vol.		3.0B	25.6M	32.1M	38.0M	32.6M

The average volatility of the S&P 500 is low, as the index value is an average of 500 stocks. The WSB portfolio stock's average daily volatility increases when there is activity on WSB, especially when there are buy signals. In total, the WSB portfolio's 21 stocks unsurprisingly generate a much lower trading volume on average than the S&P 500. Still, an average stock listed in the S&P 500 corresponds to a daily trading volume of c. 5.98 million USD (1/500 of the S&P 500 volume), i.e., the average WSB portfolio stock is traded with a five times higher volume than the average S&P 500 stock. The average daily trading volume is significantly higher on days with WSB activity, especially if a buy signal is detected (c. 49% higher than on average). This implies that WSB activity increases when markets are volatile and volume is high or possibly that increased WSB activity may lead to higher volatility and volume in financial markets.

5.3. Reliability of Buy and Sell Signals

The average performance of the portfolio has proven beneficial, but which percentage of signals from WallStreetBets was actually successful? For each of the WSB portfolio's tickers, we evaluate the accuracy of the community's buy and sell signals by determining the respective price developments over specific time windows – if the price increased after a buy signal, it is considered successful (conversely for sell signals). This assessment can reveal whether the buy and sell signals would have led to better investment decisions than investing continuously or randomly in the same stocks over the same time frame, with a focus on short-term windows. We compare the success rate of all buy signals for a given stock to three baselines of when a stock is

bought or sold on: *equally distributed days*, *randomly distributed days* (average over 5 independent trials for a more robust estimate), *every day*. For the former two, we use a sample size identical to the number of buy signals. These baselines simulate three alternative types of hypothetical investors: one investing every $\lfloor \frac{D}{n} \rfloor$ days, one choosing n random days to invest, and one investing every day (with n being equal to the number of buy or sell signals that the baseline is compared to and D being the total number of days in the data).

Our analysis reports the average price change after predefined time windows depending on when an investment was made. The difference between three different baseline distributions are not particularly strong. Buy signals on WSB portfolio stocks were accurate (positive price development): after one day in 51.75% of the cases, after three days in 52.88%, after one week in 55.02%, after one month in 61.28%, and after three months in 69.94% of cases. In comparison, the baselines achieved accuracies of 51.08–52.80%, 52.70–53.19%, 54.40–54.93%, 58.12–59.19%, and 60.52–61.60% respectively. Within the shorter time frames of one day to one week, following the buy signals is approximately equally often successful as the baselines. However, WSB's buy signals seem to be more successful across longer time frames, as the buy signals' success rate is up to 5.4% higher (after one month) and up to 15.6% higher (after three months) than the baselines' success rates. This means that an investment decision following a WSB buy signal increased in price after three months in almost 70% of the cases.

Table 4. WSB portfolio's price development for different investment patterns (Jan 2019 – Apr 2021)

Investment Pattern	Avg. price change (%) after			
	1 day	1 week	1 month	3 months
Average	0.17	1.24	5.14	16.65
Mention	0.29	2.07	7.77	27.32
Buy Signal	0.29	2.19	9.88	32.03
Sell Signal	0.65	3.77	8.12	28.53
Equal Dist.	0.19	1.21	5.10	14.64
Random Dist.	0.13	1.44	5.36	15.86

In the next step, we review the average price changes to produce more detailed insights than comparing general upward or downward movement of prices. Table 4 compares the average price development for different investment patterns: the equally and randomly distributed baselines introduced above, and WSB activity signals. The results suggest that investment decisions based on WSB activity entailed significantly larger subsequent price increases. Again, it should be noted that the detected sell signals have been highly unsuccessful, as the prices increased even more

after those days than when selling on days without community activity. For a more fine-granular analysis, the comparison of the price development of buy signals versus the average may be reviewed on a per-stock basis. Table 5 provides notable examples of the price development for specific WSB portfolio stocks for investments following a buy signal as well as the average price development. Such a fine-grained analysis shows that following buy signals only led to a consistently better outcome for a few selected stocks: \$GME, \$AMC, \$BB, \$AMZN, \$TLRY, \$TSLA – the first three (\$GME, \$AMC, \$BB) belong to the four *meme stocks* that were at the center of attention at the end of January 2021, \$TLRY is a cannabis stock that achieved *meme stock* status in February due to its sudden increase in popularity, and \$TSLA has been a highly popular stock throughout the last years. However, WSB did not produce reliable buy signals for the other prominent *meme stocks* \$NOK and \$ACB.

Table 5. Average price change vs. average price change after buy signal (selected WSB portfolio stocks; January 2019 – April 2021)

Ticker	Pattern	Avg. price change (%) after			
		1 day	1 week	1 month	3 months
GME	Average	0.76	7.00	29.95	109.16
	Buy Signal	2.86	22.35	100.66	355.36
AMC	Average	0.40	2.30	8.70	8.80
	Buy Signal	0.92	11.50	33.75	28.11
BB	Average	0.10	0.73	3.68	10.43
	Buy Signal	0.10	1.12	4.54	14.38
NOK	Average	0.00	-0.06	-0.52	-2.29
	Buy Signal	-0.28	-0.64	-3.00	-0.69
ACB	Average	-0.04	-0.24	-2.73	-10.25
	Buy Signal	0.29	-1.39	-4.45	-9.38
AMD	Average	0.22	1.48	6.19	19.61
	Buy Signal	0.09	1.35	5.34	17.10
AMZN	Average	0.10	0.68	2.74	9.12
	Buy Signal	-0.01	0.77	3.19	11.80
BA	Average	0.03	0.34	0.80	-1.56
	Buy Signal	0.08	0.03	-0.58	0.46
NIO	Average	0.36	2.76	13.60	62.35
	Buy Signal	0.02	0.36	12.30	69.92
TLRY	Average	0.08	0.80	2.02	7.54
	Buy Signal	0.46	0.45	14.12	35.62
TSLA	Average	0.35	2.54	11.95	45.19
	Buy Signal	0.55	3.49	13.31	47.18

When focusing on price developments after three months, a few tickers can be added to the list above: \$AAPL, \$AMZN, \$BA, \$FB, \$GE, \$NIO, \$QQQ – following buy signals with these stocks resulted in higher growth than the average price development. This means that despite volatility in the short term, following

buy signals would have turned profitable after three months for 14 of the 20 stocks. When excluding the stocks that incurred losses on average (\$NOK, \$ACB, \$BA), following buy signals achieved a 56% (mean) or 17% (median) higher price growth than the average price increase. In conclusion, we distinguish two cases: In the short term (1 day, 1 week), an investment strategy following buy signals actually was not more successful than the average price increase with respect to short-term windows since January 2019 – instead the success of buy signals was mostly driven by *meme stocks*, which biased the average results (especially GameStop). In the longer term (3 months), however, the buy signals turned out to provide better growth than the average price development on average. This confirms the findings of the previous analysis: trusting buy signals on the very short term seems to be successful as often as it fails, but for more patient investors, the portfolio of WSB's preferred stocks performed well and the buy signals turned out to be valuable indicators.

5.4. Proactive vs. Reactive Signals

While WSB has shown a tendency to bet on stocks that were likely to perform better than the broader market, an investment strategy following WSB buy signals would not necessarily be successful. With the goal of separating valuable signals from those that are unsuccessful, we investigate whether signals can be classified as proactive or reactive. Reactive signals (e.g., a buy signal just after the stock price rose significantly) are usually not helpful (at least in the short term) – they can lead to financial losses if the timing is poor, e.g., when GameStop was briefly valued above \$300. However, they occur very frequently due to users sharing their gains or losses from recent investments or posting news articles on specific stocks that recently made headlines due to notable price changes. The most valuable form of (short-term) investment advice is one received before a change in stock price occurs, thus constituting a proactive signal. In order to validate this assumption, we have labelled the data for each case (with x referring to 1 day, 3 days, or 1 week).

Reactive: the price increase in the preceding x is *higher* than the price change in the following x .

Proactive: the price change in the preceding x is *lower* than the price increase in the following x .

Following this classification, only 46.5% of all buy signals were proactive on average. Our analysis reveals clearly different patterns for reactive and proactive signals – investments following reactive signals performed significantly worse in comparison. This suggests that a distinction between both types

may be useful to identify the most valuable pieces of investment advice. Nonetheless, even the average of reactive signals achieved higher growth than the average price increase. Investments made after reactive buy signals follow after a period of above-average price increase and show above-average volatility and trading volume on days of the signal (conversely for proactive buy signals). These effects are amplified when the time x for detecting a reactive or proactive signal is longer, but apply to all evaluated time windows. Drawing on this distinction, an investor following only proactive buy signals would have been able to achieve 700% higher growth after a single day and 50% higher growth after three months on average. This conclusion also applies when excluding the *meme stocks* of January 2021 from the calculation – in this case, 2,400% higher growth after a single day and 75% higher growth after three months could be achieved on average. This suggests that distinguishing reactive and proactive buy signals is highly relevant for WSB-guided investors to be successful in the short term.

Obviously, a stock's future price development is uncertain, making it much easier to decide retrospectively whether a signal was proactive or reactive. Hence, we also evaluated a simple heuristic for making this distinction using historic price information only: We only select buy signals on days that show a closing price below its moving average over 30 days. While this does not detect all proactive signals, it provides a reasonable workaround. In future work, more sophisticated analysis approaches could be investigated.

5.5. Before and After the Hype

It is important to note that the rapid increase in community members and generated content in and after January 2021 entails a strong shift in the dataset towards the newest posts and discussions, which were also coupled with high-volatility price movement in the financial markets. In order to assess whether the observations from the analysis above also hold for the time before the January hype phase, we applied the same analysis to the subset of data from the time frame spanning January 1, 2019 through December 31, 2020. The results indicate that for the time frame before 2021, buy signals were much less effective over the short term than in 2021, but still showed better performance over longer time periods than the average price development. The overall impact of buy signals is weaker in contrast to the post-hype values – nonetheless, similar tendencies can be detected. We obtain the same effect for the comparison of reactive and proactive signals.

6. Conclusion

WallStreetBets has come to prominence as a forum for unconventional high-risk investment discussions. This paper assesses the reliability of investment advice encountered in the subreddit. The analysis reveals that recent activity on WSB could have served as profitable investment advice on multiple levels:

1. WSB's discussions have focused mostly on sectors that performed better over the reviewed time frames than the S&P 500 index. Investing in these sectors and holding for a longer period would have yielded better results than the average market growth.
2. A hypothetical portfolio containing the most frequently and consistently discussed stocks outperformed the S&P 500 on average over the reviewed time windows.
3. An investment strategy following buy signals would have achieved higher average growth on the longer term than distributing investments equally or randomly over the same time window – while in the short term these strategies performed similarly.
4. Distinguishing signals as proactive or reactive helps in achieving even more successful outcomes on average as well as in the short term. Moving averages constitute a simple heuristic for this distinction.
5. The observed tendencies were amplified by WSB's strong growth and the success of *meme stocks* in 2021, but they can still be found in a weaker form when only considering earlier data, corroborating the above results.

Overall, these results suggest that an investment-focused social community such as WallStreetBets can indeed serve as a source of valuable investment advice. Although care needs to be taken to identify promising signals among the posts, the community constitutes a freely accessible alternative to other sources such as financial news outlets or individual experts on Twitter. While retail traders can leverage WSB as a source of financial knowledge and entertainment, professional investors may benefit from regular visits by learning about trending topics and stocks as well.

References

- Agrawal, P., Buz, T., & de Melo, G. (2022). WallStreetBets beyond GameStop, YOLOs, and the moon: The unique traits of Reddit's finance communities. *AMCIS 2022 Proceedings*, 8.
- Anand, A., & Pathak, J. (2021). WallStreetBets against Wall Street: The role of Reddit in the GameStop short squeeze. *IIM Bangalore Research Paper*, (644).

- Baumgartner, J., Zannettou, S., Keegan, B., Squire, M., & Blackburn, J. (2020). The pushshift reddit dataset. *Proceedings of the International AAAI Conference on Web and Social Media*, 14.
- Boylston, C., Palacios, B., Tassev, P., & Bruckman, A. (2021). WallStreetBets: Positions or Ban [arXiv:2101.12110v1].
- Bradley, D., Hanousek Jr, J., Jame, R., & Xiao, Z. (2021). Place Your Bets? The Market Consequences of Investment Advice on Reddit's Wallstreetbets [SSRN 3806065].
- Chohan, U. W. (2021). Counter-Hegemonic Finance: The Gamestop Short Squeeze [SSRN 3775127].
- Daas, P. J., & Puts, M. J. (2014). *Social media sentiment and consumer confidence* (tech. rep.). ECB Statistics Paper.
- Duz Tan, S., & Tas, O. (2021). Social Media Sentiment in International Stock Returns and we Trading Activity. *Journal of Behavioral Finance*, 22(2), 221–234.
- Gorodnichenko, Y., Pham, T., & Talavera, O. (2018). *Social media, sentiment and public opinions: Evidence from #Brexit and #USElection* (tech. rep.). National Bureau of Economic Research.
- Hajli, M. N. (2014). A study of the impact of social media on consumers. *International Journal of Market Research*, 56(3), 387–404.
- Hasso, T., Müller, D., Pelster, M., & Warkulat, S. (2022). Who participated in the GameStop frenzy? Evidence from brokerage accounts. *Finance Research Letters*, 45, 102140.
- Howard, P. N., Duffy, A., Freelon, D., Hussain, M. M., Mari, W., & Maziad, M. (2011). Opening closed regimes: What was the role of social media during the Arab Spring? [SSRN 2595096].
- Jones, C. M., Reed, A. V., & Waller, W. (2021). When Brokerages Restrict Retail Investors, Does the Game Stop? *Columbia Business School Research Paper*.
- Jones, H., & Hietanen, J. (2022). The r/wallstreetbets 'war machine': Explicating dynamics of consumer resistance and capture. *Marketing Theory*, 14705931221114172.
- Kim, K.-S., Sin, S.-C. J., & Yoo-Lee, E. Y. (2014). Undergraduates' use of social media as information sources. *College & Research Libraries*, 75(4), 442–457.
- Leimeister, J. M. (2010). Collective intelligence. *Business & Information Systems Engineering*, 2(4), 245–248.
- Lyócsa, Š., Baumöhl, E., & Vÿrost, T. (2022). YOLO trading: Riding with the herd during the GameStop episode. *Finance Research Letters*, 46, 102359.
- Manikonda, L., Beigi, G., Kambhampati, S., & Liu, H. (2018). #Metoo through the lens of social media. *SBP-BRIMS*, 104–110.
- Mendoza-Denton, N. (2021). "Sticking It to the Man": r/wallstreetbets, Generational Masculinity and Revenge in Narratives of our Dystopian Capitalist Age. *Anthropology Now*, 13(1), 91–99.
- Nguyen, T. H., Shirai, K., & Velcin, J. (2015). Sentiment analysis on social media for stock movement prediction. *Expert Systems with Applications*, 42(24), 9603–9611.
- Preece, J., & Shneiderman, B. (2009). The reader-to-leader framework: Motivating technology-mediated social participation. *AIS Transactions on Human-Computer Interaction*, 1(1), 13–32.
- Schou, P. K., Bucher, E., Waldkirch, M., & Grünwald, E. (2022). We did start the fire: R/wallstreetbets, 'flash movements' and the Gamestop short-squeeze. *Academy of Management Proceedings*, 2022(1), 14028.
- Stieglitz, S., & Dang-Xuan, L. (2013). Emotions and information diffusion in social media — Sentiment of microblogs and sharing behavior. *Journal of Management Information Systems*, 29(4), 217–248.
- Sul, H. K., Dennis, A. R., & Yuan, L. (2017). Trading on Twitter: Using social media sentiment to predict stock returns. *Decision Sciences*, 48(3), 454–488.
- Tajudeen, F. P., Jaafar, N. I., & Sulaiman, A. (2016). Role of social media on information accessibility. *Pacific Asia Journal of the Association for Information Systems*, 8(4), 3.
- Umar, Z., Gubareva, M., Yousaf, I., & Ali, S. (2021). A tale of company fundamentals vs sentiment driven pricing: The case of GameStop. *Journal of Behavioral and Experimental Finance*, 30, 100501.
- Wang, C., & Luo, B. (2021). Predicting \$GME Stock Price Movement Using Sentiment from Reddit r/wallstreetbets. *Proceedings of the Third Workshop on Financial Technology and Natural Language Processing*, 22–30.
- Yu, Y., Duan, W., & Cao, Q. (2013). The impact of social and conventional media on firm equity value: A sentiment analysis approach. *Decision Support Systems*, 55(4), 919–926.