Analysis of Social Media Driven 'Meme Stocks'

Context

This project is largely inspired by the rise of 'meme stocks' in January of 2021, when predictions about an impending increase in the price of Gamestop caused the stock to see an unprecedented increase in value. This high profile event spawned a number of other 'meme stocks' which are stocks given a lot of attention by small investors in forums like wallstreetbets and appear to move erratically. We sought to answer the question "Can wallstreetbets influence the movement of stocks?" and "Can we base a predictive model on wallstreetbets activity?"

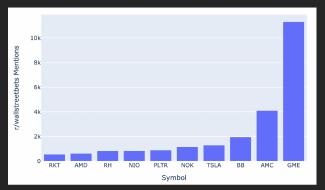
ENGINEERING HONORS PROGRAM
UNIVERSITY OF MICHIGAN

With the Mentorship of Prof. Daniel Romero

Findings & Discussion

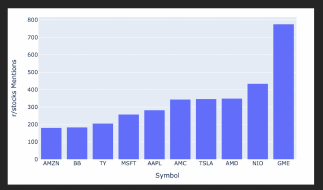
To discern what stocks were popular on wallstreetbets and how that differs from other investing-focused subreddits, we compared stocks mentioned on r/wallstreetbets to those on r/stocks. We also what words r/wallstreetbets used in 2020 to what it used in 2021 to explore the changes in topics of discussion. Comparing r/wallstreetbets activity to GME and AMC stock prices was somewhat promising, so we fit a linear regression model on the words appearing in posts in a day on r/wallstreetbets from January 2021 to September 2021 to predict the change in closing price between that day and the end of the next day. We then used this to forecast the price of GME from October 2021 to December 2021 using 1-day lags. Using derivative instruments, we can derive the implied volatility of GME, a metric calculated from the price of derivatives, independent of the direction of price. We then refit a linear model onto the implied volatility data, and forecasted the 1-day changes. Due to the protected nature of financial data, we were only able to find daily GME information.

Top Tickers /r/WallStreetBets



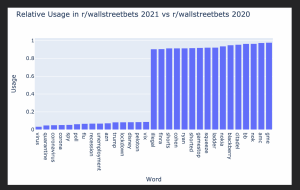
r/wallstreetbets tended to focus intensely on a few stocks, namely GME, AMC, and BB
Top tickers stocks...

Top Tickers /r/Stocks



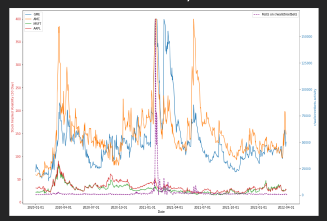
r/stocks focused on a lot of the same stocks as r/wallstreetbets, however it was much more balanced, and also focused on some well established "blue chip" stocks like AAPL and MSFT.

Relative Word Usage



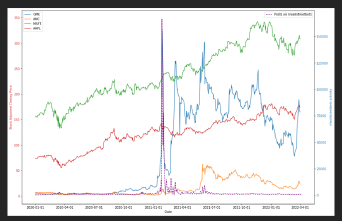
Changes in vocabulary indicate a shifting focus from coronavirus-related issues in 2020 to meme stocks and the market manipulation narrative.

WSB Activity vs IV



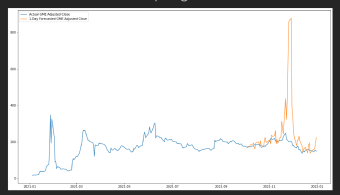
The stocks popular on r/wallstreetbets are smaller and have higher IV to begin with. It also appears that the spikes in volatility among stocks popular on r/wallstreetbets coincide with the spikes in activity.

WSB Activity vs Price



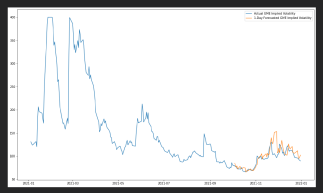
The explosion of activity on r/wallstreetbets coincides with the increase in price in the stocks that are popular on it, even though many popular stocks were relatively unaffected.

Forecasted Price
1-Day Lag



Changes in vocabulary indicate a shifting focus from coronavirus-related issues in 2020 to meme stocks and the market manipulation narrative.

Forecasted IV
1-Day Lag



Our regression model for IV was also unable to significantly predict changes in GME's IV, however it was much more promising than the model for price.

Conclusions

Findings

Based on our research, it's hard to conclude anything specific about how r/wallstreetbets influences stocks. However, it seems that the discussion of these stocks on r/wallstreetbets tends to coincide with a surge in price in these stocks, and with it a surge in implied volatility. It seems likely that there is a relationship between the two sources. Despite our models not being very predictive, the stock market is a complex adaptive system that should correct for any predictive relationship quickly, and there may have been relationships at the time. The implied volatility prediction was much more successful than predicting the change in price, suggesting that r/wallstreetbets may be a good source of whether something will change in the price of a stock in the future.

Future Direction

There are many exciting directions this project could move in. For one, a major problem we faced was the unavailability of market data. Because statistical arbitrage opportunities close quickly, it follows that working at shorter time intervals should reveal more of the relationship (assuming a predictive relationship exists). Another direction is to use more complicated models to analyze r/wallstreetbets' posts on the market, and to look at some of the many spin-off subreddits formed in the aftermath of Gamestop. There were a number of factors that our linear model couldn't consider that are probably significant. We worked with a fixed time interval, but there's probably interesting information to find in variable time intervals. Specifically, we would like to research the conditions that cause a normal stock to become a meme stock, and whether you can predict what stock will become the next meme stock. Lastly, in comparing the differences between r/stocks and r/wallstreetbets, we discovered major differences in vocabulary that highlight the differences in these communities. We think we could use different subreddits to train models to predict the subject of a document and the formality of different subreddits.

Nicholas Anason > Engineering Honors

Alexander Daines > Cognitive Science Honors