Analyzing Price Fluctuations in Reddit's 'Meme' Stocks

Jeff Hollis

Advisor: Professor Kornhauser

Submitted in partial fulfillment

of the requirements for the degree of

Bachelor of Science in Engineering

Department of Operations Research and Financial Engineering

Princeton University

June 2022

I hereby declare that I am the sole author of this thesis.

I authorize Princeton University to lend this thesis to other institutions or individuals for the purpose of scholarly research.

_____

Jeff Hollis

I further authorize Princeton University to reproduce this thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

_____

Jeff Hollis

Abstract

Reddit's so called *meme stocks* have taken over the stock market over the course of the 2021 calendar year. The events that transpired surrounding these stocks have created phenomenon that did not exist in the past. Fueled by millions of people following the r/wallstreetbets subreddit, these highly volatile stocks have seen more than 1000% increases in just a several days solely due to viral internet popularity. High jumps in the prices of these stocks are rooted in social sentiment towards those companies rather than the company's true economic value. The power of the ever evolving internet and the collection action of a large group of everyday retail investors sparked these large price fluctuations. This thesis takes a broad look at exactly how much of a connection r/wallstreetbets actually had to the movement in stock price and stock volume over the course of the meme stock craze and beyond.

Acknowledgements

Thanks to the friends who have made the Princeton experience one of a kind, it was an honor.

Contents

Chapter 1

Introduction

In the start of 2021, we have observed a large number of retail investors relying almost exclusively on social media indicators originating from places like Twitter and Reddit. This sparked an entirely new phenomenon where a large number of these retail investors jumped on the bandwagon and invested in these select stocks, regardless of what the fundamentals of the stock say. These stocks present a high risk, high reward opportunity—wrong timing can lead to massive loss.

The most notable and largest meme stock surge was the GameStop surge in late January. GameStop (GME) started 2020 trading at $6 per share, but dropped to as low as $3 per share. GME experienced high levels of volatility in 2020 and eventually rose to a price of $20 per share as Chewy, Inc. co-founder Ryan Cohen disclosed an investment in GameStop. GameStop had already been receiving attention from Reddit in 2019 and 2020. Discussion of the so-called "short-squeeze" had already been underway. At the end of 2020, short interest in GameStop was hovering around 100%, mainly from large institutional investors believing GameStop was fundamentally overpriced as the era of physical copies of games dwindled. This short interest inspired a frenzy that originated from subreddit r/wallstreetbets to buy shares of GME. The number of traders trading GME skyrocketed in January 2021, and conse-

quently so did the price. Many institutions were forced to buy shares of their own to cover their massive short positions as the price of GME "went to the moon", reaching its high of $483.00 per share on January 28, 2021. GME reached as high as a 140% increase in price in a single trading day.

GameStop isn't the only company that has undergone this treatment though. Other prominent examples include companies like AMC, Palantir, and Virgin Galactic.

This paper aims to evaluate how linked r/wallstreetbets is to the price and volume movements of stocks in hopes to uncover how strong the connection is between r/wallstreetbets and meme stocks.

1.1 r/wallstreetbets

r/wallstreetbets has been the central hub for casually discussing the stock market for some number of years now. This place of discussion has been an area of collaboration for casual retail traders who are trying to improve their portfolio. Additionally, it's a forum board for those who wish to see unfiltered opinions from a wide range of people. This aspect is appealing to many, as main stream media will often give a seemingly surface level description on a given topic.

Because of Reddit's score based structure on posts, many basic users of WSB are able to see the consensus best posts without any effort. This gives room for people who just want to see the most popular content for a couple of minutes, while also leaving room for the more experienced users who enjoy taking a deeper dive into the content. This kind of mechanism makes it easy for new users to enjoy the forum without putting much effort into finding the "best posts".

1.1.1 The GameStop Frenzy

Both WSB and GameStop popularity began steeply rising around the same time in early January 2021. Tempers from the WSB community rose as the short interest in GameStop became larger, and finally the WSB community was seemingly able to break through and surge GameStop's share price up to levels it had never seen before. With GameStop's sharp rise in price, WSB's popularity rose right along side it (1.1).

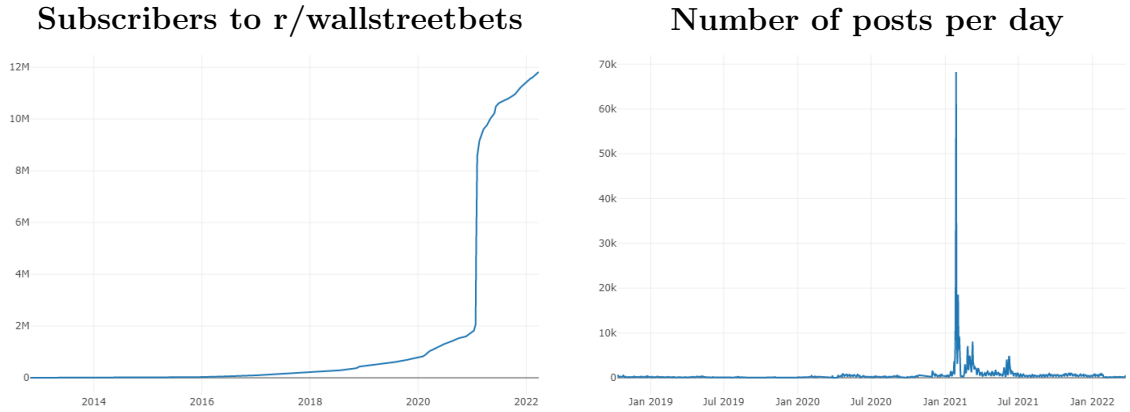

Figure 1.1

The numbers of subscribers to r/wallstreetbets rose from 2 million to 10 million in less than a month after its nearly 10 year existence.

1.1.2 Posts

WSB has a variety of the types of content that is posted to it. Some posts give in relatively deep financial analysis, while many others are simply humorous threads made as jokes. The variety in the content is quite immense, (1.2) and (1.3) show some examples of posts on WSB in the context of GameStop.

As you can see, the exact topic of the post can vary significantly. One of the posts is a picture showing someone giving away GameStop products to a local hospital, while another posts shows a billboard purchased with the a message signalling to people that GME is "going to the moon". (1.3) depicts the topic of gain/loss on

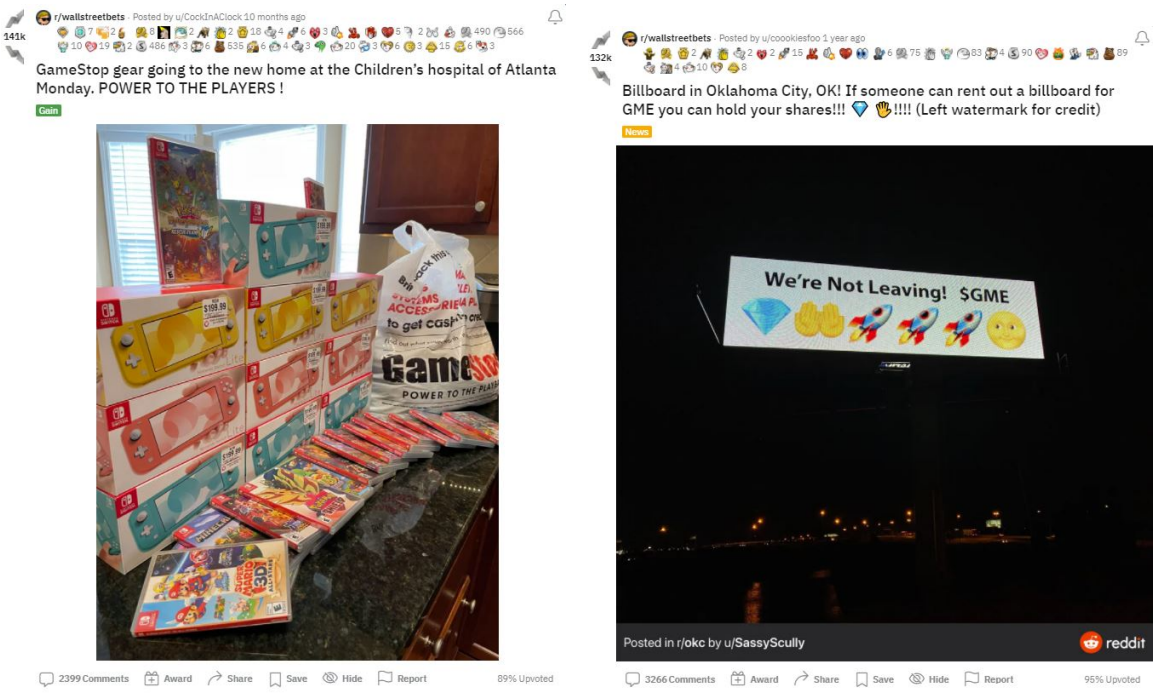

Figure 1.2: Typical GameStop posts on r/wallstreetbets.

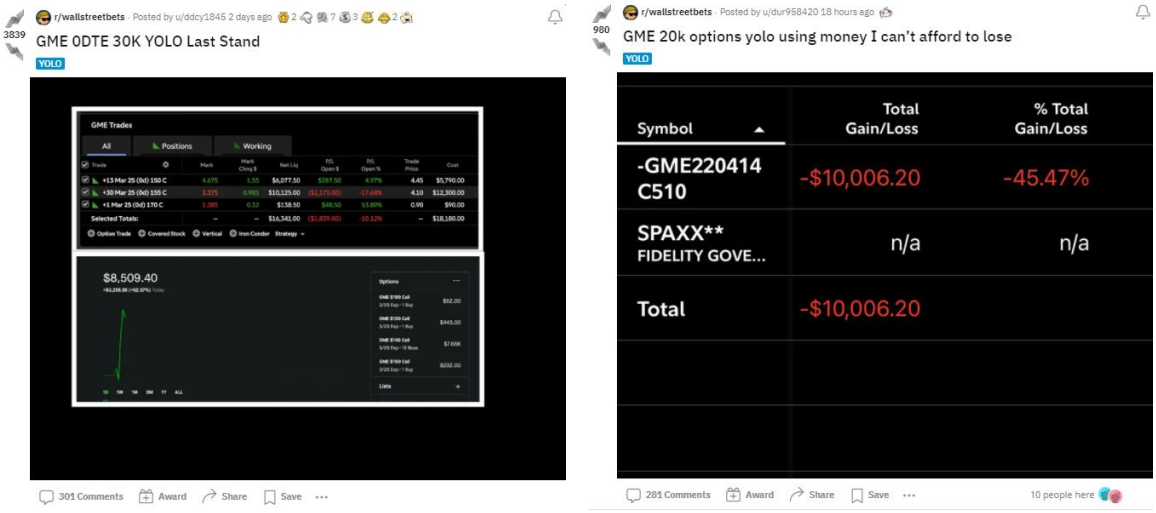

Figure 1.3: Many posts concerning meme stocks are related to sharing the incredible loss or gain that your gamble on the stock has garnered. These are two examples of gain/loss posts for GameStop.

WSB. Some of the most popular posts of all time on the subreddit are screenshots of profit/loss sections of a person's portfolio, showing either massive loss of massive gain on heavily gambled positions.

Notice the "post score" on the top left of each post in Figure 1.2, totalling 141,000

and 132,000 respectively. The number of comments can be seen in the bottom left of each post as well. These numbers will become relevant in the analysis section. Notably, the number of comments for the post on the right is significantly higher than the number of comments for the post on the left, even though the left post has a higher post score.

1.2 Robinhood

Robinhood is a key player in the history of meme stocks. Being the most popular commission free trading platform available, Robinhood stands at the forefront of the discussion of meme stocks. The "average Joe" in the stock trading world is likely to use Robinhood as their platform to make trades.

Notably, Robinhood halted the purchase of stocks for several publicly traded companies including GameStop, Blackberry, AMC, and others. As the r/wallstreetbets short-squeeze of GME occurred at the end of January, many large investment funds who owned a short position on Gamestop lost billions. One of those funds was ' "Melvin Capital Management", owned by the parent company "Citadel, LLC", gave Robinhood about 40% of their revenue. This caused major outrage with Robinhood from all retail investors, which ended up fueling the fire even more after the hault was uplifted. Robinhood and Citadel underwent investigation called on by Congress and entered into a lawsuit that was filed on behalf of the retail investors who weren't able to sell during the halt and consequently lost money. This lawsuit was eventually dismissed, allowing Robinhood to get off scot free[1].

Chapter 2

Literature Review

2.1 On the "mementum" of Meme Stocks

The most prominent scholarly work done on meme stocks is a piece of work titled "On the 'mementum' of Meme Stocks" [5]. This paper is an initial attempt at identifying "meme" periods of a stock's life "based on regime-switching cointegration and a simple procedure to identify them from market and social data." The technique described uses twitter data and stock data to determine if and when any specific stock undergoes 'mementum'.

Costola, Iacopini, and Santagiustina employed a specific procedure for identifying what they describe as a "meme period". They firstly provide a formal definition for their term "mementum" and create a characterization from the pairwise cointegration of 1) price and social media presence, and 2) volume and social media presence. Next, based off of the characterization provided from the pairwise cointegration, they use a regime-switching cointegration model to ultimately identify the meme periods of the stock.

This framework is an entirely data-driven approach which leads to zero bias for identifying meme periods. There could be some bias when hand selecting stocks as

meme stocks, so making these decisions solely from the data provided is the least misleading approach to take.

2.1.1 Social Media Presence

There are many different routes one could venture on when trying to decide what type of data to scrape from social media with regard to meme stocks. In this piece, Costola, Iacopini, and Santagiustina chose to go onto the largest social media platform in the world to get their data: Twitter. Specifically, they decided to use all tweets that contained an image because of the definition of a *meme*. In addition, they restrict the tweet to contain at "least one emoji symbol related to the meme stock booming", as seen in (2.1).

2.1.2 Formal definition of "Mementum"

Costola, Iacopini, and Santagiustina create a definition of their concept mementum that requires three conditions to hold.

Condition 1 requires that there needs to be cointegration present the price/tweet and volume/tweet pairs at the same exact time. Condition 2 requires that both the price/tweet and volume/tweet cointegration pairs need to enter the period of cointegration at the same moment. Condition 3 requires that the cointegration from condition 1 is present for at least some predetermined amount of time. If these three conditions are satisfied, then the stock has undergone *mementum*.

2.1.3 Results

Figure 2.2 shows the results from the initial mementum study.

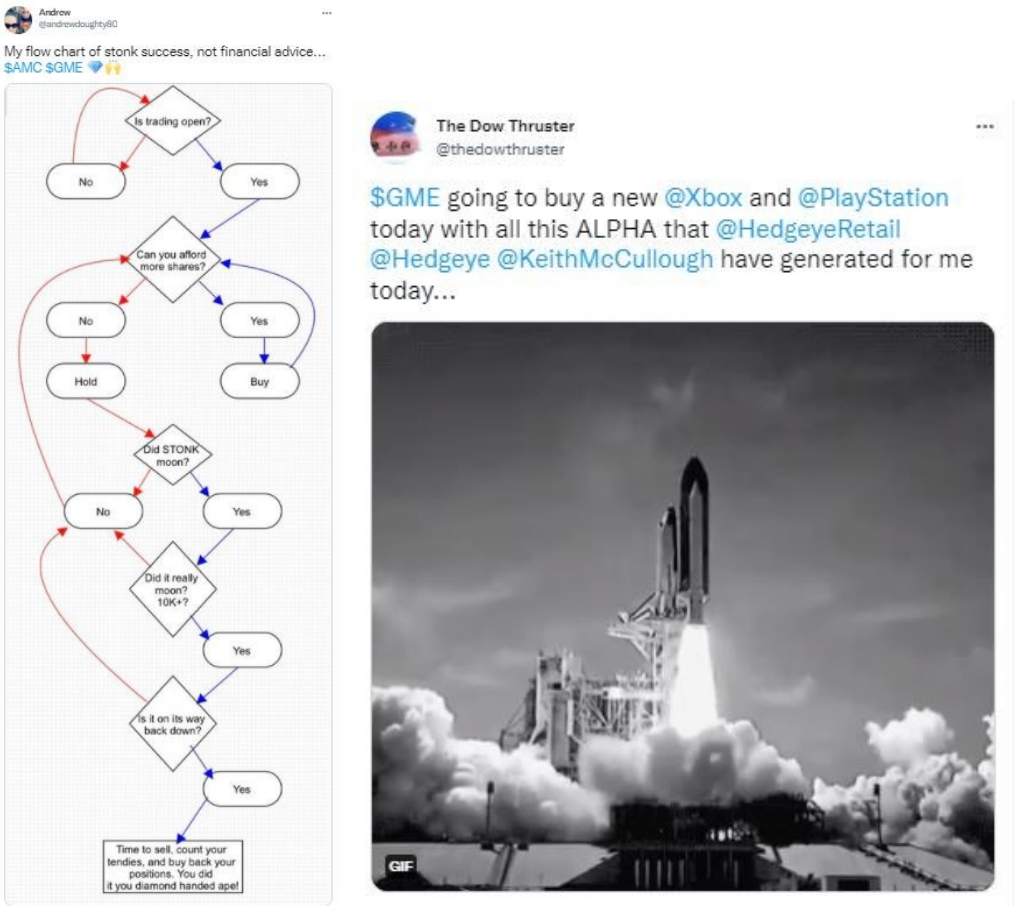

Figure 2.1: These are two examples of meme stock related tweets provided from *mementum*. The left tweet mentions GME and AMC with an image saying to hold with "diamond hands". The right tweet mentions GME with a GIF of a rocket launching from earth representing the price of GME. Left tweet link: https://twitter.com/andrewdoughty80/status/1381818643789479936. Right tweet link: https://twitter.com/thedowthruster/status/1349398282842353666

2.2 On the Efficiency of Meme Stocks

Another piece of work surrounding the topic of meme stocks employs market efficiency tests to test the market efficiency of meme stocks relative to the overall market over the course of the start of Covid-19 until mid 2021[4]. They provide evidence that meme stocks have always been weakly efficient (prices are not predictable, they follow a random walk). Interestingly, several indicators give the result that the S&P500 was not weakly efficient during the beginning span of the Covid-19 crisis.

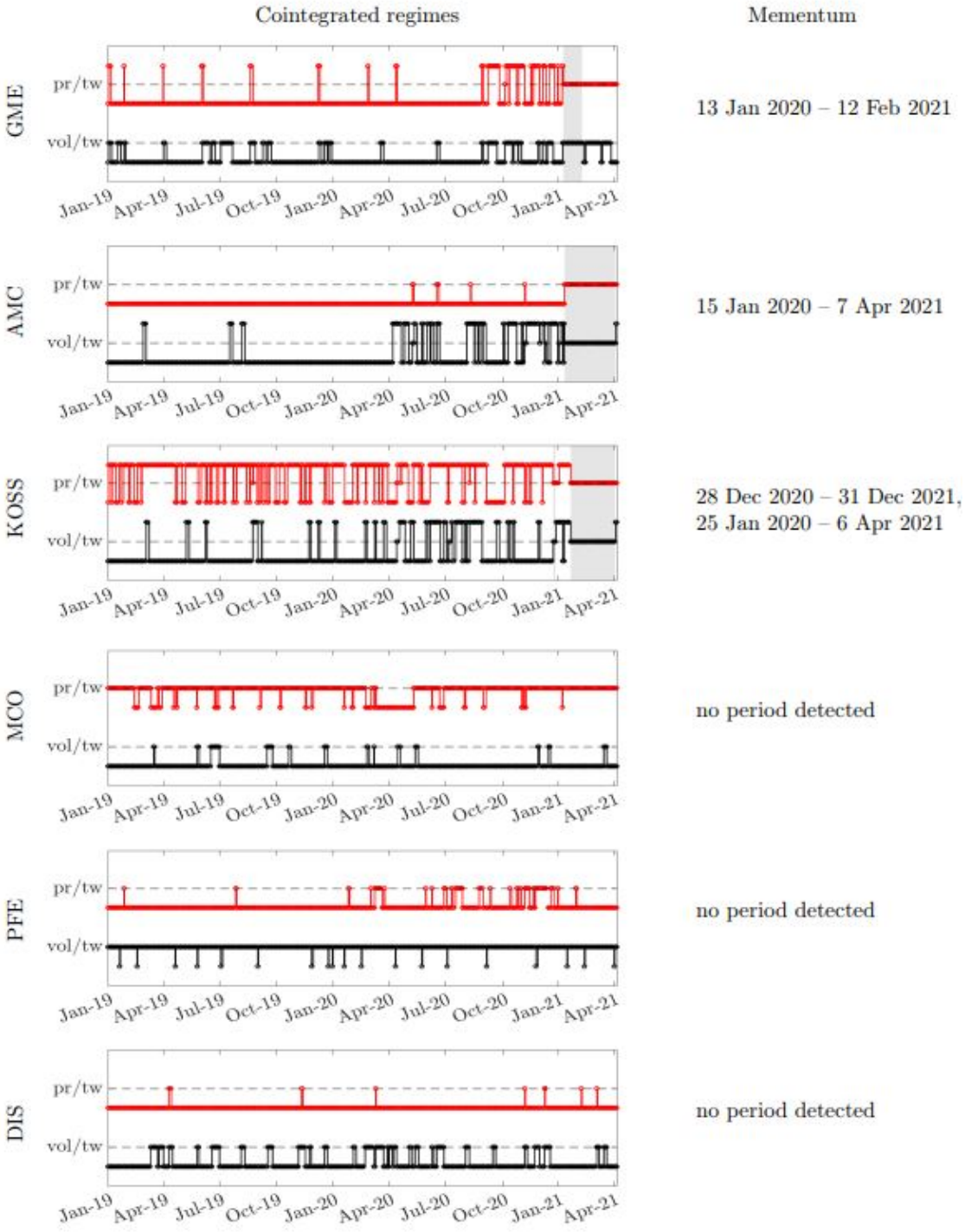

Figure 2.2: The dashed line represents the cointegration regime in each time series. The gray shaded area is a period of mementum.

Chapter 3

Data

3.1 Stocks

To collect stock data, I used the python yfinance API[2].

3.2 Reddit

Collecting data from reddit can be slightly complicated. PRAW is the main Reddit API used for scraping and extracting data from Reddit. Unfortunately there are some limitations from this API that can't be escaped. Namely, there is no capability to obtain any historical data from any posts or subreddits using PRAW. Of course for this project, the historical post data is precisely what is needed.

Fortunately, there is a much more convenient API called Pushshift[3] for accessing Reddit's historical data. The pushshift.io Reddit API was designed and created by the /r/datasets mod team to facilitate further function beyond that of what comes with the PRAW API.

Chapter 4

Analysis of Problem

4.1 Preliminary Analysis

The meme stock surge initially occurred in late January 2021. This is when the most prominent meme stock, GameStop, had its biggest spike. Shortly after this point in time other stocks entered the world of r/wallstreetbets and transformed into meme stocks themselves, like AMC ($AMC), Tilray ($TLRY), and Nokia ($NOK). All stocks have their own unique reason for rising to prominence in the wallstreetbets subreddit. GameStop was a nostalgic millennial company riding on thin valuations and receiving heavy short interest. Wallstreetbets users didn't like the fact that their childhood safe haven was looking more and more like it going to disappear, so they rallied together in an attempt to save it. Nokia, a dinosaur of a phone company famous for its brick-like phones shares a similar story. Like GameStop, the number of shorted shares for Nokia increased significantly in the beginning of 2021 which, in turn, infuriated the r/wallstreetbets community.

In this analysis of meme stocks, I will be using market data in the form of closing price and volume traded per day in combination with several types data points that show trends in the popularity and discussion of specific stocks in r/wallstreetbets. I

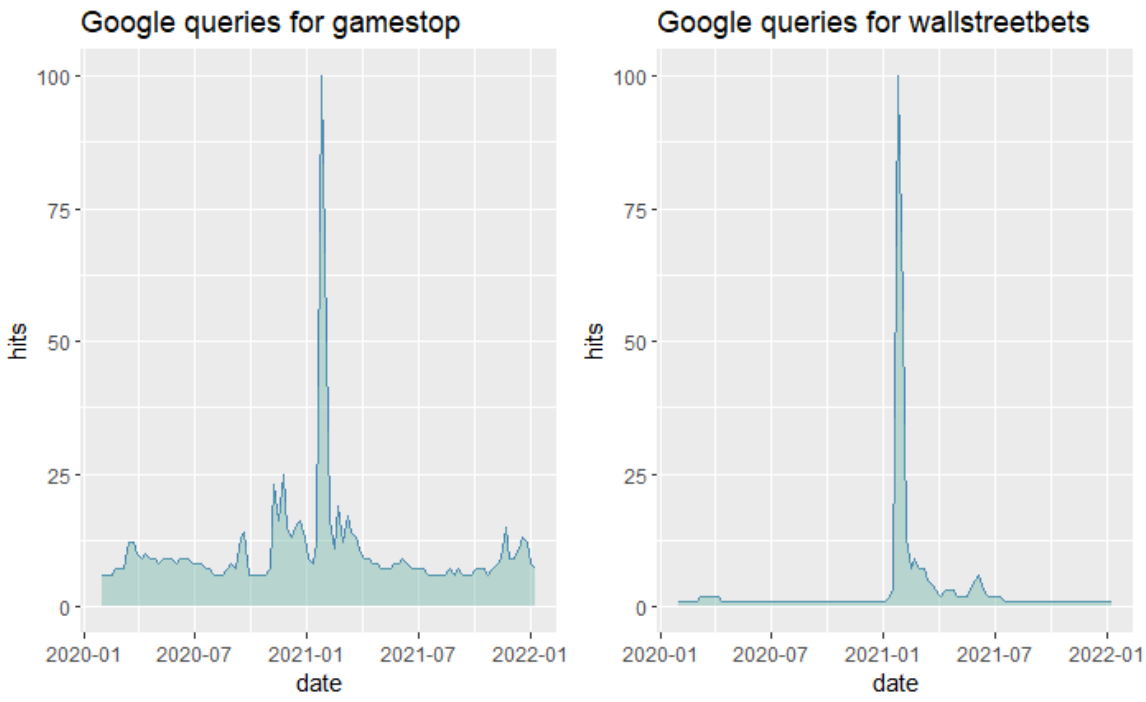

Figure 4.1: Charts showing the spike in google searches over time. Obtained using R package gtrendsR.

decided to capture data each trading day in the time period of February 1st, 2020 until January 11, 2022. In total, this amounts to 490 active trading days, with the meme stock surge occurring in the middle of the time frame.

The data I chose to collect from r/wallstreetbets come in the following forms: number of new posts containing the stock's ticker symbol in the title, cumulative reddit 'score' (the reddit score equates to the number of upvotes minus the number of downvotes on a post) on posts that contain the stock's ticker symbol in the title, and the number of cumulative comments on posts that contain the stock's ticker symbol in the title. Each of these data were collected on every trading day for the duration of the time period I mentioned previously.

As you can see in Figure 4.2, there is a sharp spike at the end of January 2021 in the number of reddit posts for each stock ticker. This spike can be similarly seen in the other categories of reddit data I collected: score and number of comments.

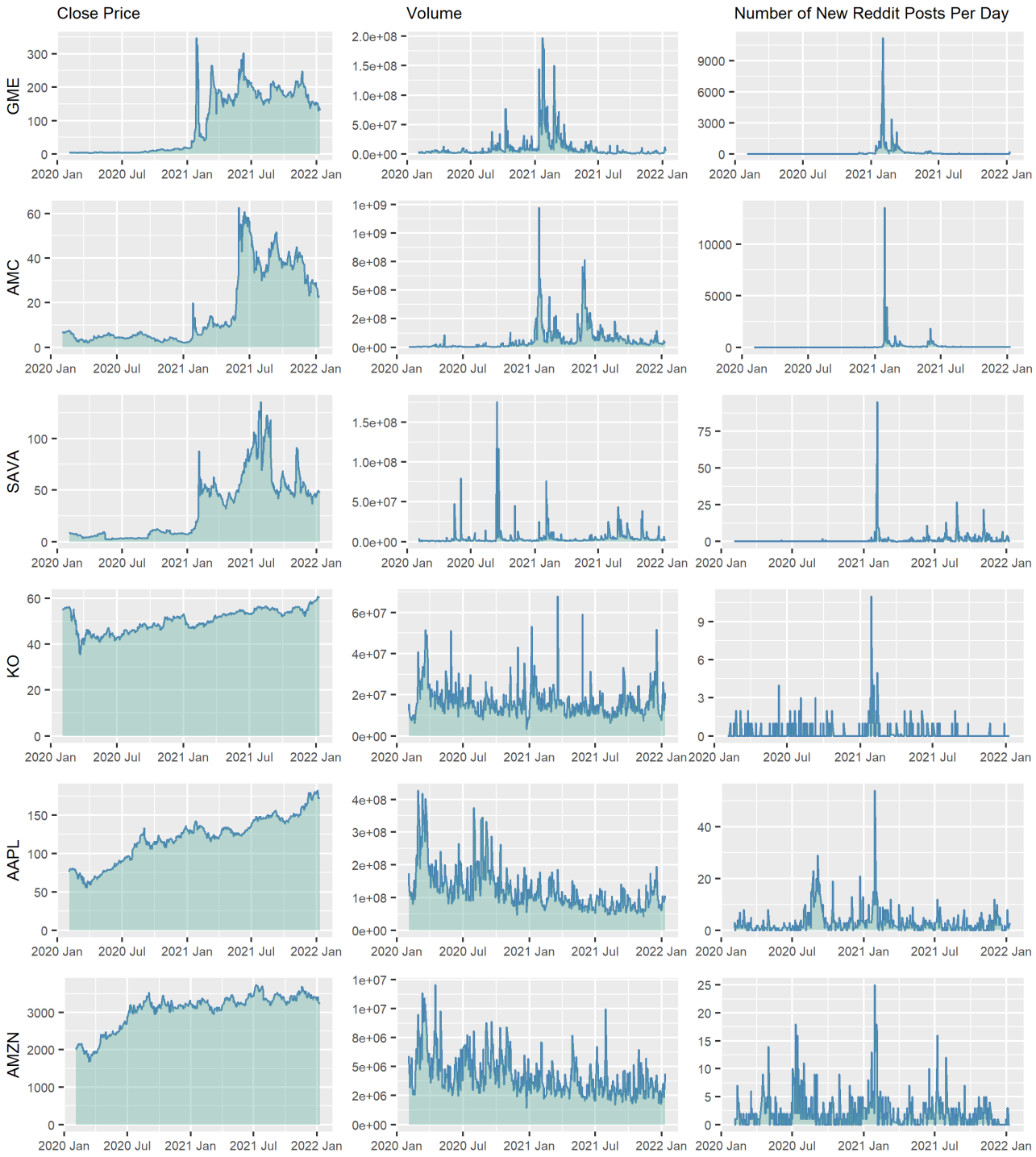

Figure 4.2: Time series plots for closing price (column 1), trade volume (column 2), and number of reddit posts (column 3) for GameStop (GME), AMC Entertainment (AMC), Cassava Sciences Inc. (SAVA), Coca-Cola (KO), Apple (AAPL), and Amazon (AMZN).

Amazingly, GameStop and AMC both were posted on WSB thousands of times per day at their peaks, with AMC's peak being the largest with over 12,500 posts in a single day. In fact, to include the peaks in the (4.2) charts, the y-scale increased so significantly that most of the rest of the chart appears to be 0. Even though each stock in Figure 4.2 appeared to have had a spike around this time, the magnitude of the spikes in the non meme stocks were significantly less than the magnitude of the spikes in the meme stocks, especially compared to GameStop and AMC. Even though it looks like each stock in the chart had similar spikes in posts around late January, it's important to notice the scale of the y-axis in each plot to see the vast difference.

Peak WSB Posts

Ticker	Date	Posts
AAPL	2021-01-28	54
AMC	2021-01-28	13526
AMZN	2021-01-28	25
CVNA	2020-04-14	5
DISCB	2021-03-31	17
GME	2021-02-01	11225
KO	2021-01-28	11
MRNA	2020-12-01	20
NOK	2021-01-27	3510
NVDA	2021-11-04	43
PLTR	2020-11-27	996
SAVA	2021-02-04	95
SPCE	2021-01-28	274
TSLA	2020-02-05	223

GameStop's highest closing price occurred on January 27, 2021, where it closed at a price of $347.51. The highest post frequency in WSB though, occurred not on January 27, but on February 1. Several other stocks though, including AMC, peaked in posts on January 28, a seemingly more fitting day for a high volume of posts given the historic GameStop price fluctuations occurring on the day prior. Even though this surge of posts on the subreddit centered around January 28, 2021, we can clearly see that a large number of stocks didn't follow this trend. Tesla peaked about a year prior in daily posts, while Moderna and Palantir both peaked in late 2020 for reasons of their own. Still though, stocks with seemingly no connection to the GameStop surge, like Coca-Cola (KO), surged in posts on January 28. The reason for this goes to the major media attention

r/wallstreetbets was getting around this time. GameStop had already been rising quite quickly for about a week at that point, garnering it major worldwide attention and in turn drawing attention to r/wallstreetbets. When GameStop had its biggest surge on January 27, it boosted WSB activity more than any other single day. This, in turn, caused the number of posts in general to go up, not just the posts for GameStop.

To get a better understanding of the overall activity inside of WSB, look at the logarithmic transformations (4.3) of the post numbers. It's clear the the activity of GameStop and AMC spiked in early 2021, and then tapered off, but the activity has not fully gone away like it did in Apple for example. In Apple's plot, there is very little activity overall, and even though there was a spike on January 28, it seemingly did not make any lasting effects on its popularity in r/wallstreetbets in the long run as it did with GameStop and AMC.

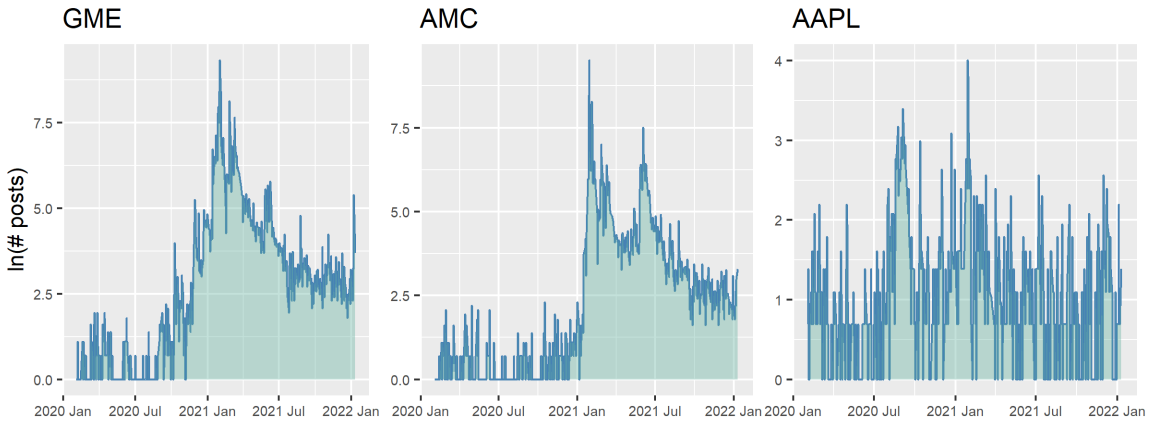

Figure 4.3: Logarithmic transformations of the post frequency data in r/wallstreetbets for stocks GameStop, AMC, and Apple.

In this analysis, I will be focusing mainly on a selection of around 13 stocks. Some of the stocks I chose with the knowledge of their existence as meme stocks, like GameStop and AMC for example. Others on the list, like Apple and Coca-Cola, I chose as controls to examine how large the differences are between the stocks that we think of as 'meme' stocks and those that are just 'normal' stocks.

Daily Returns

Feb 3, 2020 - Jan 10, 2022

Ticker	Mean	Variance
AMC	0.01100	0.03000
GME	0.01400	0.01800
SAVA	0.01100	0.01800
SPCE	0.00130	0.00460
MRNA	0.00670	0.00340
CVNA	0.00310	0.00320
PLTR	0.00280	0.00240
TSLA	0.00500	0.00220
NOK	0.00160	0.00120
NVDA	0.00370	0.00110
AAPL	0.00200	0.00057
AMZN	0.00120	0.00041
KO	0.00034	0.00029

Figure 4.4: List of daily returns which includes mean and variance. The stocks that are highlighted blue are those that are generally considered to be 'meme' stocks. List is sorted by largest variance to smallest variance.

Figure 4.4 shows the selection of stocks and their daily returns and variance for the period of time under study. Not surprisingly, the meme stocks are at the top of the list (sorted by decreasing variance). Not only did they score higher variances than the non meme stocks, but their returns tend to higher as well. Notably, even though the most famous meme stock is GameStop, AMC actually tops GameStop in volatility. GameStop had the highest single day spike, but AMC is actually significantly more volatile than any other stock on the list, including GameStop.

4.2 OLS Regression

4.2.1 Price

A preliminary test that can be conducted to determine some of the relationships involved in stock movement and reddit activity is a linear regression. This test will show early results for whether price or volume movement is explained by reddit activity or not. Moreover, the p-values for the features in the regression will be good indications for which predictors are superior. Although these results might appear somewhat convincing, more tests need to be run in order to tell the full story about these relationships. For a given stock k, I will use P_t^k as the closing price on day t, F_t^k as the number of new WSB posts on day t, C_t^k as the number of new comments on new WSB posts on day t, and S_t^k as the score of new WSB posts on day t. The model is as follows:

$$P_t^k = \beta_0 + \beta_1 F_t^k + \beta_2 C_t^k + \beta_3 S_t^k + \epsilon_t$$

I will be using log values for the testing. In financial models, logarithmic transformations tend to be the norm, especially when it comes to prices[17].

The results from Figure 4.5 are actually incredibly close to what was expected. The stock with the highest R^2 was GameStop, followed by AMC, and then Virgin Galactic Holdings ($SPCE). In fact, from the selection of stocks in the list, the preselected 'meme stocks' were indeed the top ones on the list, while the stocks that were thought of as non-meme stocks, are sitting at the bottom. The only exception is Nokia ($NOK). Nokia is certainly deemed a meme-stock, but sits with a shockingly low R^2 value to what would be expected. Nokia's price though, didn't follow a trajectory like GameStop or AMC. Instead of having a major spike and then diminishing thereafter, it had a minor spike (relative to GameStop's spike) and continued to rise throughout the year as seen in Figure 4.6.

OLS Regression Results for Price				
		p-value		
Ticker	R.Squared	# Posts	# Comments	Post Score
GME	0.583	0.000	0.175	0.000
AMC	0.414	0.000	0.000	0.000
SPCE	0.323	0.000	0.000	0.038
SAVA	0.273	0.000	0.047	0.003
PLTR	0.221	0.000	0.003	0.912
NVDA	0.101	0.000	0.143	0.545
MRNA	0.059	0.000	0.825	0.012
AAPL	0.044	0.000	0.290	0.052
NOK	0.026	0.055	0.149	0.035
KO	0.025	0.384	0.171	0.800
AMZN	0.019	0.107	0.146	0.004
CVNA	0.014	0.416	0.213	0.216
TSLA	0.014	0.017	0.941	0.101

Figure 4.5: These are the results from ordinary least squares regression with model $P_t^k = \beta_0 + \beta_1 F_t^k + \beta_2 C_t^k + \beta_3 S_t^k$, where P_t is the closing price, F_t is the number of posts, C_t is the number of comments, and S_t is the cumulative score of the posts on day t for stock k. The values in each feature column are the p-values associated with those features for the given stock. The R.Squared column is the R^2 associated with the model for a given stock.

Notably, the 'number of comments' predictor generally seemed to do a poor job in explaining the close price. In fact, among all stocks, only AMC (AMC), Virgin Galactic (SPCE), Cassava Sciences (SAVA), and Palantir (PLTR) regard the number of comments to be an important predictor of the price. The number of posts seems to be the best overall predictor for closing price. Almost every stock deemed the number of posts important at a 0.05 significance level, and certainly any stock with a reasonable R^2 deem it as a reliable predictor. Similarly, the post score appears to be a reasonably good predictor in most cases. The top four stocks by R^2 consider it

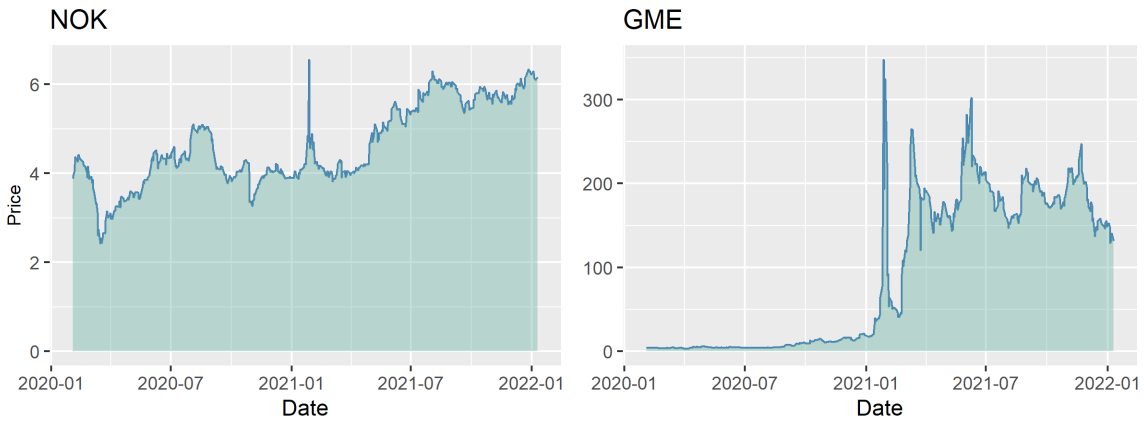

Figure 4.6: Comparison of Nokia and GME stock prices.

to be an important predictor.

4.2.2 Volume

Now volume will be regressed on instead of price, using OLS. V_t^k is the volume traded on day t for stock k. Each feature variable is the same as in the price regression. The model is as follow:

$$V_t^k = \beta_0 + \beta_1 F_t^k + \beta_2 C_t^k + \beta_3 S_t^k + \epsilon_t$$

The results for the regression on volume are slightly different than the results for the regression on price. AMC has the highest $R^2 = 0.78$. This is higher than any R^2 found in the price regression by quite a bit. Most stocks with high R^2 values for the price regression also have high R^2 values for the volume regression, but there are notable differences in the results. For one, Nokia has risen dramatically in R^2. According to the regression, the r/wallstreetbets data explained almost nothing about the price movements for Nokia, but for Nokia volume movement on the other hand, the Nokia WSB data explains about 40% of the movement. Tesla and Moderna had similarly steep R^2 rises in the volume regression. Overall, the average R^2 rose from

OLS Regression Results for Volume				
		p-value		
Ticker	R.Squared	# Posts	# Comments	Post Score
AMC	0.779	0.000	0.027	0.005
SPCE	0.475	0.000	0.434	0.780
NOK	0.415	0.000	0.568	0.180
PLTR	0.414	0.000	0.491	0.625
GME	0.367	0.233	0.929	0.000
MRNA	0.367	0.000	0.000	0.282
SAVA	0.309	0.000	0.660	0.676
TSLA	0.192	0.000	0.824	0.337
AMZN	0.094	0.387	0.563	0.000
AAPL	0.071	0.057	0.786	0.130
NVDA	0.061	0.044	0.771	0.436
CVNA	0.057	0.000	0.424	0.934
KO	0.004	0.209	0.767	0.169

Figure 4.7: These are the results from ordinary least squares regression with model $V_t^k = \beta_0 + \beta_1 F_t^k + \beta_2 C_t^k + \beta_3 S_t^k + \epsilon_t$, where V_t is the closing price, F_t is the number of posts, C_t is the number of comments, and S_t is the cumulative score of the posts on day t for stock k. The values in each feature columns are the p-values associated with those features for the given stock. The R.Squared column is the R^2 associated with the model for a given stock.

about 0.16 average in the price regression to about 0.28 in the volume regression.

Once again, number of comments is the weakest predictor. Only two of the stocks on the list, AMC and Moderna, deemed number of comments as a significant predictor. Post score is also a relatively weak predictor with only three stocks in the list deeming it as important at a 0.05 significance level. Number of posts, like in the price regression, is the strongest predictor by quite a big margin. The p-value for number of posts is 0 for most stocks, with the exception of GameStop. The regression results find that GameStop's volume movement is only explained by post score, an attribute

unique to only itself and Amazon.

4.3 Spurious Regressions in Time Series

One problem that can occur in time series analysis is the problem of spurious regressions, or spurious correlations. Spurious regressions occur in time series when there is evidence of a linear relationship between variables which are non-stationary when those variables have no relation to each other in actuality. Granger and Newbold (1974) were the first to elaborate on this topic, showing that many econometric series are integrated (hence non-stationary), and that these data that are used in a regression that find a high R^2 value are likely to be independent of each other[9]. The regression residuals[1] are important in discovering spurious regressions. Namely, if the residuals are shown to be autocorrelated there is strong evidence of a spurious regression.

4.3.1 Stationarity

A random process $\{X_t\}$ is considered to be stationary if its statistical properties do not change as time progresses. This means that properties like the mean and variance do not change from one time index to the next.

There are several methods available to empirically test stationarity, one of which being the Dickey-Fuller test[6] which was later extended into the Augmented Dickey-Fuller test.

The time series in figure Figure 4.9 explain 68% of each other's movement according to an OLS model. Clearly this isn't actually correct though, as the series were generated independently of each other. Using the Augmented Dickey-Fuller test on the residuals of the regression, it is found that a rejection of the null hypothesis of

[1]The residuals ϵ_t are the error terms in the regression.

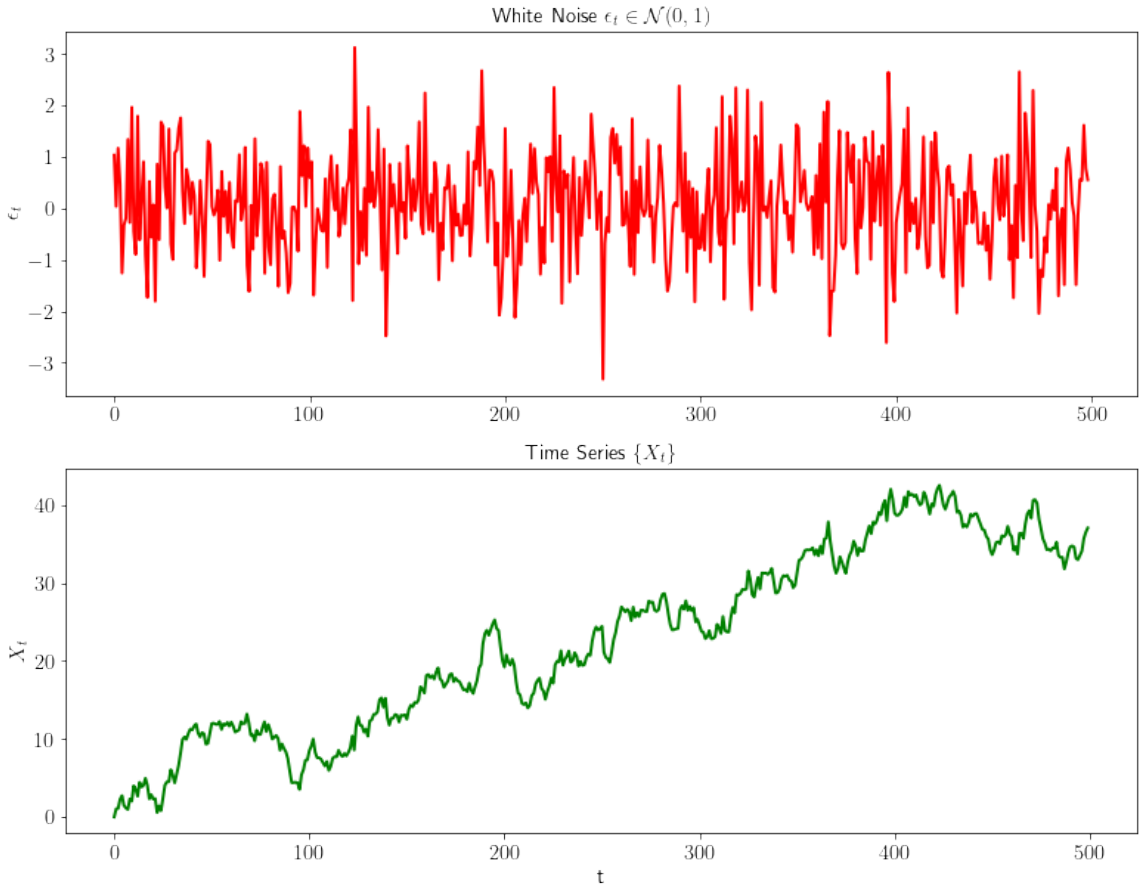

Figure 4.8: Example of a stationary series versus a non-stationary random walk.

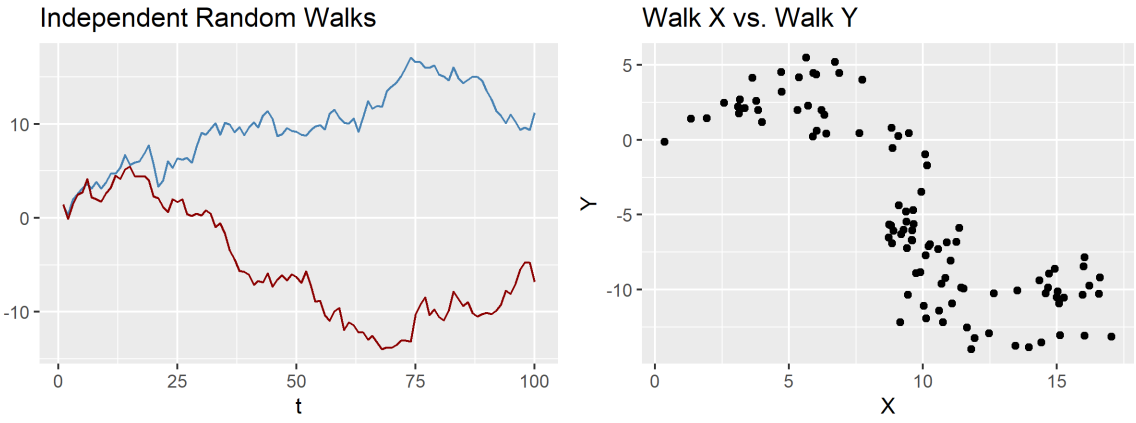

Figure 4.9: The blue and red lines are independent random walk X and Y respectively. X and Y are independent random walks with error terms $\epsilon_t \in \mathcal{N}(0,1)$. Regressing these series against each other using OLS produces an R^2 of 0.68.

a unit root being present is failed, hence the residuals are not stationary. This is evidence that a spurious regression has taken place. Alternatively, if the residuals appeared stationary, there would be evidence of cointegration of the two time series.

4.3.2 Checking for Spurious Relationships

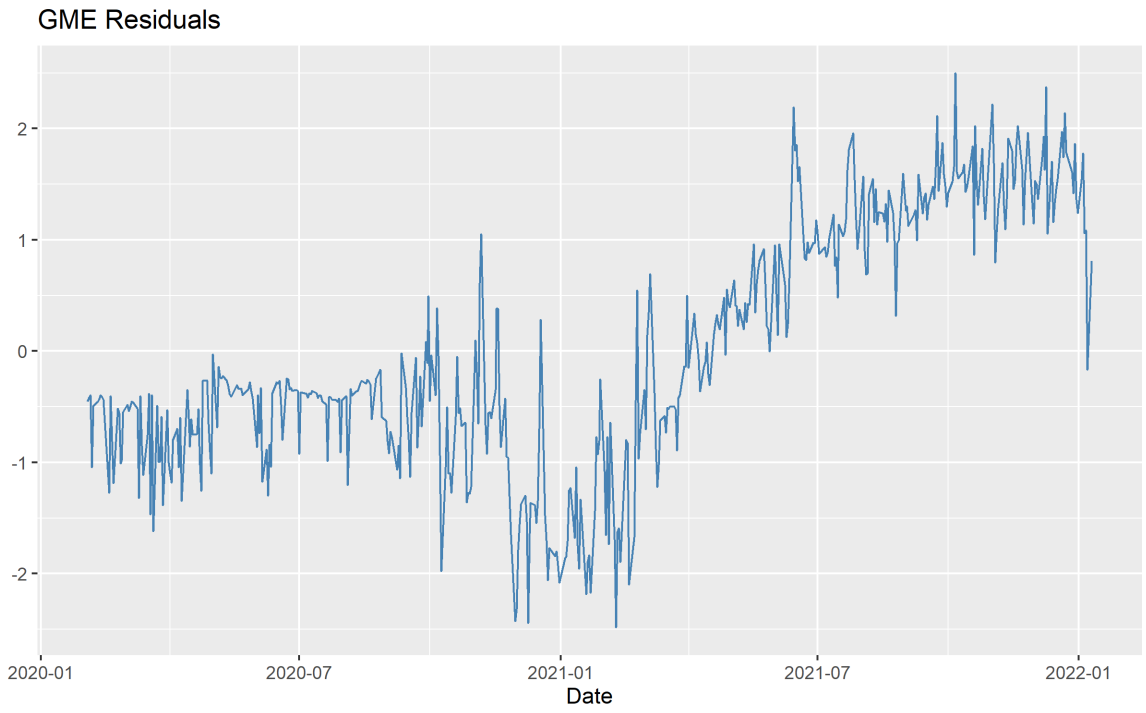

Figure 4.10: Plot of residuals from GameStop's OLS regression on price.

Naturally, the next step is to test the residuals of the prior regressions for stationarity. After taking a quick glance at GameStop's residuals (4.10), it is apparent that stationarity is not present.

Indeed, after testing stationarity in each set of residuals from the regressions on price, each regression appears to have non-stationary residuals. For the regressions with small R^2, this result doesn't mean much because those models aren't very meaningful in the first place. For the regressions that produced relevant R^2 numbers, we must view those models with caution because the results could be spurious as sug-

gested by the non-stationary residuals.

There are some interesting findings for the regressions performed on volume. Most of the stocks on the list come back with findings of non-stationarity in their residuals. Virgin Galactic (SPCE) and Palantir (PLTR), however, were found to exhibit stationarity in their residuals to a significant extent (4.11). This means that there is solid evidence for the case of a concrete (non-spurious) relationship between the reddit data and the stock volume over the entire time period for these two stocks.

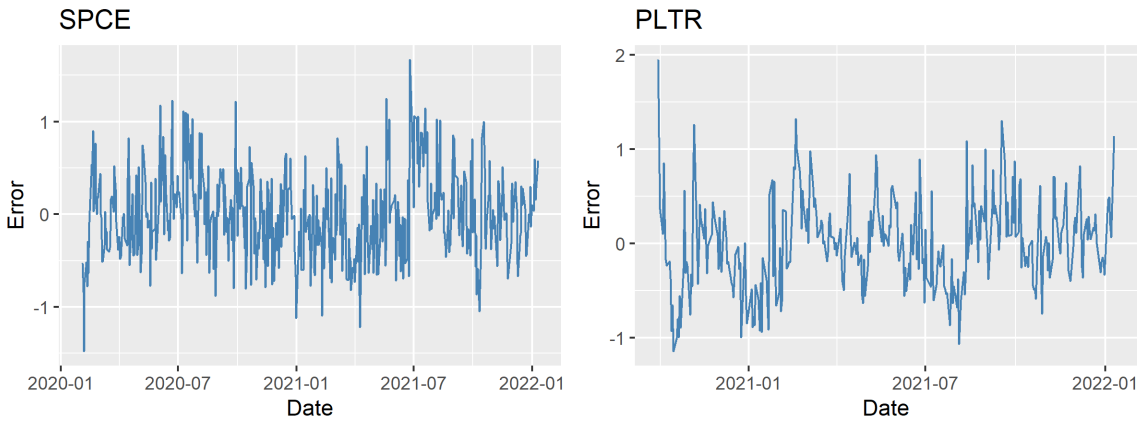

Figure 4.11: Plots of the residuals that were found stationary from the reddit data that was regressed on volume for the stocks Virgin Galactic and Palantir.

There seems to be a significant, non-spurious, connection between the number of posts, and volume for SPCE and PLTR. The p-value for the number of posts predictor was in fact the only significant predictor for both of these stocks. Of course, their R^2 values were nothing to be amazed with, 0.48 and 0.41 respectively, but they both still were near the top of the list, suggesting they are more concretely linked with reddit than almost every other stock.

It comes into question though, why are these stocks in particular the only stocks to achieve this kind of relationship with the volume throughout this time period. There seems to be a reliable explanation for why Palantir has this relationship. Palantir went public on September 30, 2020. It is the only stock in the list of stocks studied to

not have data in the entire time range I have explored (Feb 2020 - Jan 2022). I would surmise that because of the later starting date in the Palantir data, it essentially spawned into the world where r/wallstreetbets was already beginning to venture into the realm of 'meme' stocks, and thus never experienced a significant regime change like the other stocks.

4.4 Cointegration

A more concrete method of determine relation among time series is to look at the cointegration of the time series. Cointegration, coined by Engle and Granger (1987), is when a series integrated of order 0 can be created from some linear combination of two autoregressive time series[7]. One example of cointegration is akin to a drunk and her dog. A drunk woman wandering about might presumably be following a random walk. Her dog, not wanting to stray far from its owner, will stay within some proximity of the drunk[12]. The two would be existing in a cointegrating relationship because the difference of the two's distances would be stationary around some number.

Engle and Granger created a very simple two step approach for determining if two series are cointegrated. Firstly, the two series in question must be integrated of order 1. In other words, both series need to be verifiably non-stationary. Next, the series are regressed on each other using OLS in order to check if the residual series is stationary. This procedure was essentially done in the previous section already, determining that none of the residual series from the price OLS were stationary. Only two of the residual series for volume OLS were stationary, those being Virgin Galactic and Palantir. Thus, Virgin Galactic and Palantir reddit data time series are said to be *cointegrated* with their respective volume time series.

Despite the fact that two cointegrating relationships were found, the correlation coefficients involved were not very large, conveying a weak relationship between the

relevant components involved. This is an important aspect of cointegration to look at. Even though cointegration might be found, there may be a weak underlying relationship present in the first place that negates the purpose of examining the cointegration.

4.5 Time Lags

Up to now, I've only looked at same day comparisons of the data. That is, I've only been comparing the price and volume data on day t to the reddit data on day t. One might wonder though, is there any prediction that can be made with the lagged version of some series to the other? In particular, it would be interesting to look at the predictive nature of the lags of the reddit data on the closing price and volume. If it's the case that there is significant predictive value in the lags of the reddit data, then it might be possible to leverage that relationship for increased chances of profit when trading these stocks.

For this analysis, I decided to use three time lags from each variable in the regressions. The model for the price regressions is as follows for stock k:

$$\begin{aligned} P_t^k = \beta_0 + \beta_1 F_{t-1}^k + \beta_2 F_{t-2}^k + \beta_3 F_{t-3}^k + \\ \beta_4 C_{t-1}^k + \beta_5 C_{t-2}^k + \beta_6 C_{t-3}^k + \\ \beta_7 S_{t-1}^k + \beta_8 S_{t-2}^k + \beta_9 S_{t-3}^k + \epsilon_t \end{aligned}$$

Likewise, the model for the volume regressions is

Price OLS using Time Lags										
		p-value								
Ticker	R^2	F_{t-1}	F_{t-2}	F_{t-3}	C_{t-1}	C_{t-2}	C_{t-3}	S_{t-1}	S_{t-2}	S_{t-3}
GME	0.61	0.00	0.12	0.00	0.99	0.58	0.37	0.01	0.28	0.00
AMC	0.49	0.23	0.60	0.05	0.00	0.01	0.00	0.01	0.29	0.00
SPCE	0.41	0.01	0.04	0.07	0.04	0.02	0.01	0.44	0.33	0.03
SAVA	0.39	0.00	0.00	0.00	0.52	0.81	0.17	0.04	0.01	0.01
PLTR	0.25	0.02	0.22	0.02	0.09	0.17	0.07	0.63	0.99	0.65
NVDA	0.15	0.01	0.02	0.01	0.16	0.38	0.29	0.83	0.64	0.75
MRNA	0.09	0.01	0.03	0.01	0.80	0.93	0.97	0.05	0.20	0.09
AAPL	0.05	0.01	0.12	0.02	0.48	0.78	0.79	0.11	0.23	0.12
KO	0.04	0.56	0.42	0.44	0.21	0.16	0.26	0.90	0.74	0.78
AMZN	0.02	0.22	0.46	0.38	0.46	0.39	0.20	0.12	0.29	0.11
CVNA	0.02	0.59	0.65	0.81	0.09	0.06	0.19	0.12	0.08	0.12
NOK	0.02	0.30	0.27	0.56	0.40	0.41	0.16	0.17	0.24	0.22
TSLA	0.01	0.19	0.75	0.46	0.95	0.78	0.65	0.38	0.65	0.42

Figure 4.12: Results from the regressions performed on price using the lagged reddit data as predictors.

$$V_t^k = \beta_0 + \beta_1 F_{t-1}^k + \beta_2 F_{t-2}^k + \beta_3 F_{t-3}^k + \beta_4 C_{t-1}^k + \beta_5 C_{t-2}^k + \beta_6 C_{t-3}^k + \beta_7 S_{t-1}^k + \beta_8 S_{t-2}^k + \beta_9 S_{t-3}^k + \epsilon_t$$

Every stock in Figure 4.12 increased in R^2 relative to the price regression with no time lags involved. This is an important finding: the lagged reddit data has higher predictive value for the future closing price of a particular stock than the concurrent

Volume OLS using Time Lags										
		p-value								
Ticker	R^2	F_{t-1}	F_{t-2}	F_{t-3}	C_{t-1}	C_{t-2}	C_{t-3}	S_{t-1}	S_{t-2}	S_{t-3}
AMC	0.72	0.00	0.19	0.20	0.69	0.86	0.19	0.15	0.63	0.22
SPCE	0.40	0.00	0.01	0.18	0.30	0.64	0.12	0.10	0.98	0.07
GME	0.33	0.26	0.29	0.28	0.64	0.99	0.47	0.01	0.20	0.02
PLTR	0.32	0.01	0.59	0.60	0.42	0.70	0.23	0.52	0.86	0.59
NOK	0.31	0.04	0.01	0.87	0.18	0.91	0.94	0.10	0.10	0.65
SAVA	0.28	0.08	0.01	0.33	0.66	0.76	0.58	0.83	0.19	0.96
MRNA	0.26	0.03	0.50	0.97	0.00	0.03	0.38	0.39	0.36	0.47
TSLA	0.12	0.01	0.30	0.71	0.39	0.26	0.41	0.43	0.71	0.90
AMZN	0.05	0.80	0.92	0.40	0.72	0.55	0.51	0.00	0.39	0.68
AAPL	0.04	0.97	0.68	0.40	0.22	0.58	0.59	0.66	0.46	0.37
CVNA	0.03	0.02	0.45	0.77	0.27	0.13	0.36	0.71	0.44	0.31
NVDA	0.02	0.24	0.93	0.46	0.65	0.92	0.33	0.62	0.58	0.64
KO	-0.01	0.63	0.85	0.91	0.59	0.69	0.47	0.84	0.70	0.81

Figure 4.13: Results from the regressions performed on volume using the lagged reddit data as predictors.

reddit data for that stock, according to these regression results.

GameStop, the top stock in R^2 for closing price, rose 0.03 points using the lagged predictors relative to using the same day predictors—a fairly negligible increase. AMC, Virgin Galactic, and Cassava all increased a little more substantially, with the increases being .08, 0.09, and 0.12 respectively. The rest of the stocks, mostly non-meme stocks, suffer from a low R^2 that couldn't be repaired by using the lagged predictors. This makes sense though, as the previously hypothesized 'meme' stocks should be the ones with the most predictive power here. Stocks like Coca-Cola and Apple get very little attention on r/wallstreetbets relative to that of GameStop and

AMC, so the OLS results confirming that the reddit data does indeed give better predictions for these stocks is a sign that the public perception of the effect of reddit on these meme stocks might be correct.

The results from the Volume OLS using the lagged reddit predictors differs significantly from the results for the price OLS with lagged predictors. In particular, instead of the R^2 increasing when looking at the lagged values, the R^2 actually decreased. Every stock was able to predict volume better using same day reddit data rather than the lagged reddit data. GameStop had a very slight decrease in predictive value for the lagged reddit data though, with the only significant predictors being the first and third lag of the score data interestingly. In fact, GameStop was the only stock on the entire list to deem score data as a significant predictor, with the exception of Amazon.

It seems fairly random which of the predictors is significant for each stock. The one consistency in the volume OLS with time lags is that each stock generally valued one or two predictors out of the nine available, never more. The first lag of the post number data (F_{t-1}) is significant more than any other predictor. Overall though, the lags generally do a poor job of explaining volume movement, only AMC has an R^2 above 0.5. That being said, the R^2 values are relatively close to the R^2 values of the OLS with no time lags.

4.6 Regime Changes

When inspecting a time series, we often time try to look at moments in time when something about the series' model fundamentally changes. For a price series over thirty years for example, one cannot expect that price series to obey the same fundamental model over the entire duration of that thirty years. In the case of GameStop, an obvious regime switch appears to occur in early 2021. GameStop was sitting

steady at around $5 per share for a majority of 2020, yet in 2021 it transformed into an entirely new stock hovering in $100 to $200 share price range for the whole year. Moreover, the volatility experienced in 2021 compared to 2020 was a remarkable difference.

4.6.1 Markov Switching Models

For the purpose of this analysis, I will employ an approach that utilizes Markov switching models. Let's say we have a stochastic process x_t with conditional distribution

$$x_t|s_t \sim \mathcal{N}(\mu_{s_t}, \sigma_{s_t})$$

for all $t = 1, \ldots, T$. This means that if we know our current state s_t, then we know the distribution that x_t is drawing from. In real situations, we of course cannot know the exact state, but we can predict it with the data. In a Markov model, the probability to transition from state m to state n in a single time period is given by

$$\mathbb{P}(s_t = n|s_{t-1} = m) = p_{mn}$$

Using this, a transition matrix P can be constructed where the element in row i and column j is the transition probability of going from state i to state j in one period. For model with three states, the transition matrix looks like

$$P = \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{bmatrix}$$

4.6.2 GameStop and Tesla Results

The R package MSwM[15] contains an implementation for Markov Switching Models that uses maximum likelihood estimation to find the parameters for specified models. Using this, the maximum likelihood estimates for regime parameters in the data are obtained. I decided to look at the results of just two regimes in the data to reduce the complexity in the each model.

First, the log of the number of posts per day are examined in this setting. The markov switching model will find regimes with the highest likelihood means and variances for the given data. I decided to use the number of posts rather than the other two reddit data categories (number of comments and post score) here because it appears to be the most effective predictor for the prices and volumes of most of the stocks in the list. Moreover, taking the log of this data ensured that the harsh peaks in the data weren't able to take over the entire algorithm which would limit the second regime to only appear on those few days in the data with exponentially higher post numbers than the rest.

GameStop

Here are the regime results for $ln(\#\text{ posts})$ for GameStop.

```
        (Intercept)(S)    Std(S)
Model 1      0.5112349 0.6390946
Model 2      4.0074243 1.4895397

Transition probabilities:
            Regime 1    Regime 2
Regime 1 0.994461153 0.006454334
Regime 2 0.005538847 0.993545666
```

This means that every data point is assumed to be drawn from one the following distributions depending on the current regime

$$(x_t|s_t = 1) \sim \mathcal{N}(0.51, 0.63)$$

$$(x_t|s_t = 2) \sim \mathcal{N}(4.00, 1.49)$$

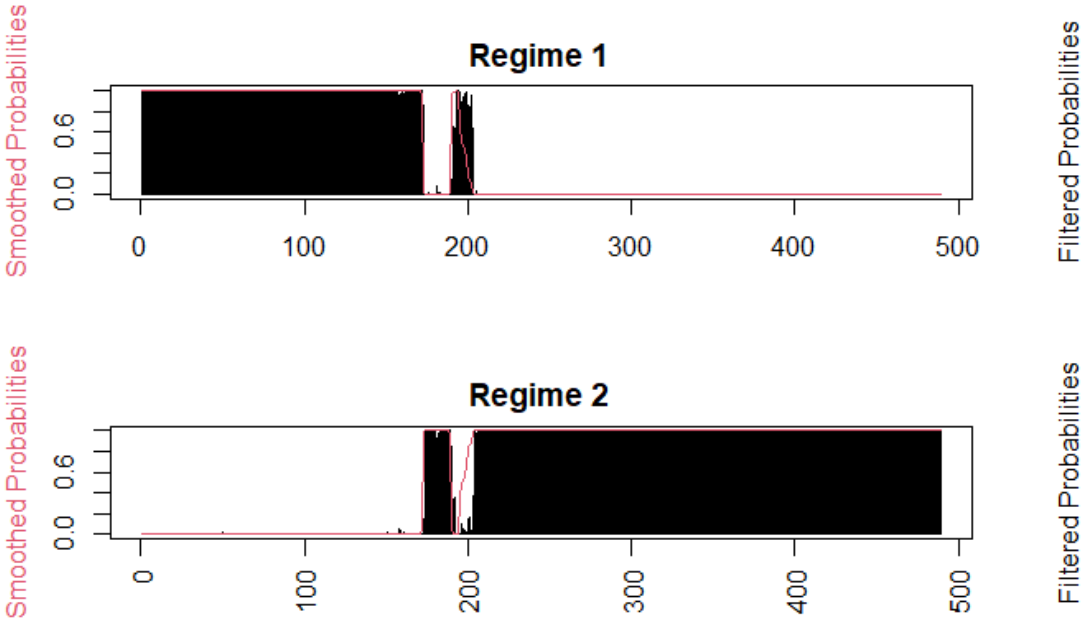

Figure 4.14: Probability of being in regime 1 or regime 2 at each point in the GameStop $ln(\#$ posts) time series.

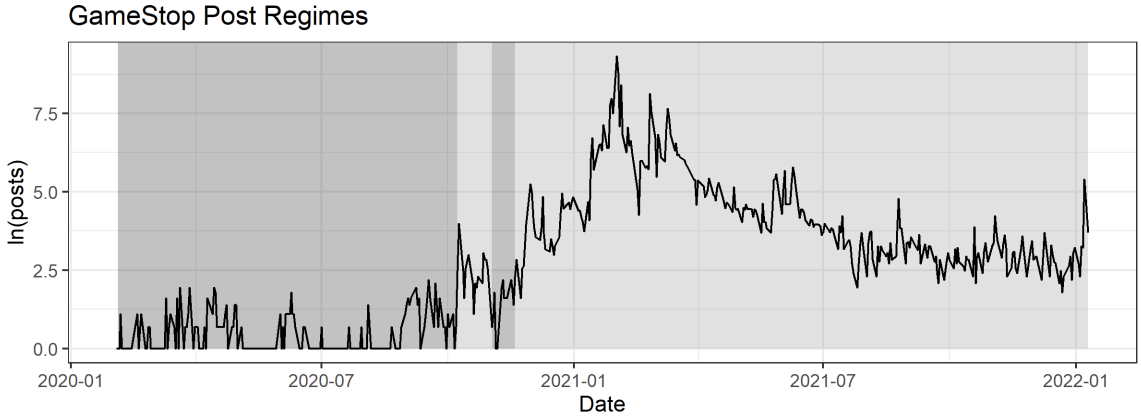

Figure 4.15: Plot of the regimes from the $ln(\#$ posts) for GameStop. The dark grey represents the time for which Regime 1 is present, and the light grey represents the time for which Regime 2 is present.

These regimes seem fairly intuitive. The posts plot in the start of the timeline

stays close to zero, and it beings to pick up around the start of 2021, and this is precisely what the regimes picked up.

Next, let's look what the estimated regimes would be for GameStop's Price time series.

```
Coefficients:
        (Intercept)(S)   Std(S)
Model 1       7.681759  4.76411
Model 2     173.173635 51.86827

Transition probabilities:
              Regime 1    Regime 2
Regime 1  1.000000e+00 0.003983569
Regime 2 3.603428e-150 0.996016431
```

The regimes found are

$$(x_t|s_t = 1) \sim \mathcal{N}(7.68, 4.76)$$

$$(x_t|s_t = 2) \sim \mathcal{N}(173.17, 51.87)$$

Not surprisingly, the second regime has nearly 11 times the volatility of the first regime, and about 22 times the mean.

The model predicts just one regime switch in totality for the entire duration of the time period. This doesn't come as much of a surprise though, as the price shifts dramatically in one specific moment and doesn't seem to return to its prior state.

The interesting aspect of looking at the regimes is to see how similar the regime switching is in both the reddit data and the price data of a specific stock. To get a numeric indicator for how similar the regime switching is for a stock, taking the proportion of days where the higher variance regimes match up in both the price and reddit data is one method. In the context of GameStop, these regimes match up in 87.7% of the time period as seen in (4.15) and (4.17). If the regime on each day was uniformly randomly chosen, we would expect for the regimes to match up around

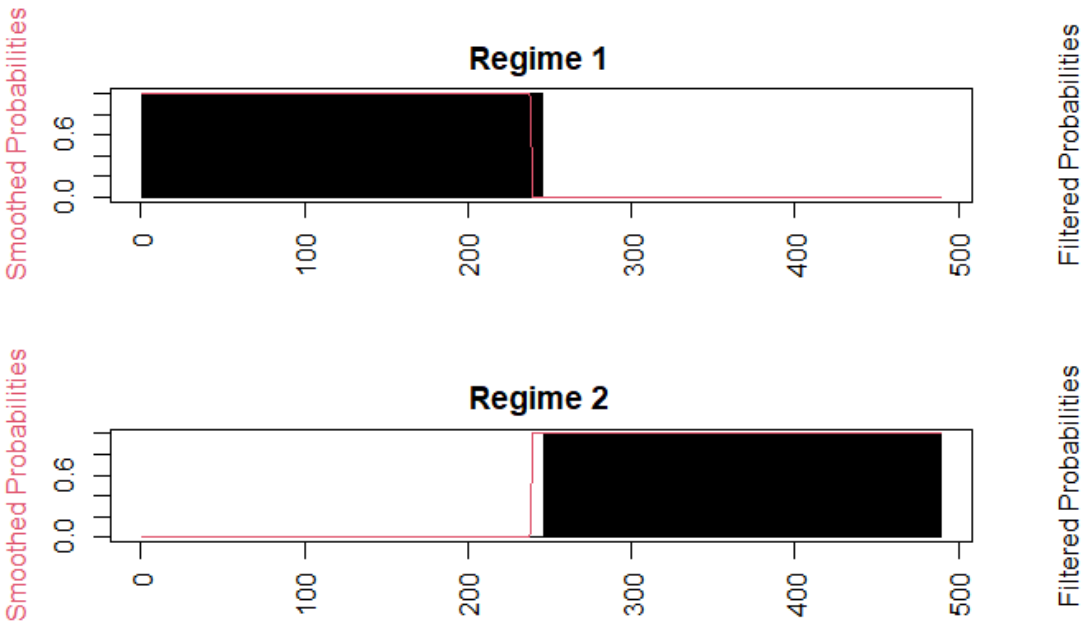

Figure 4.16: Probability of being in regime 1 or regime 2 at each point in the GameStop price time series.

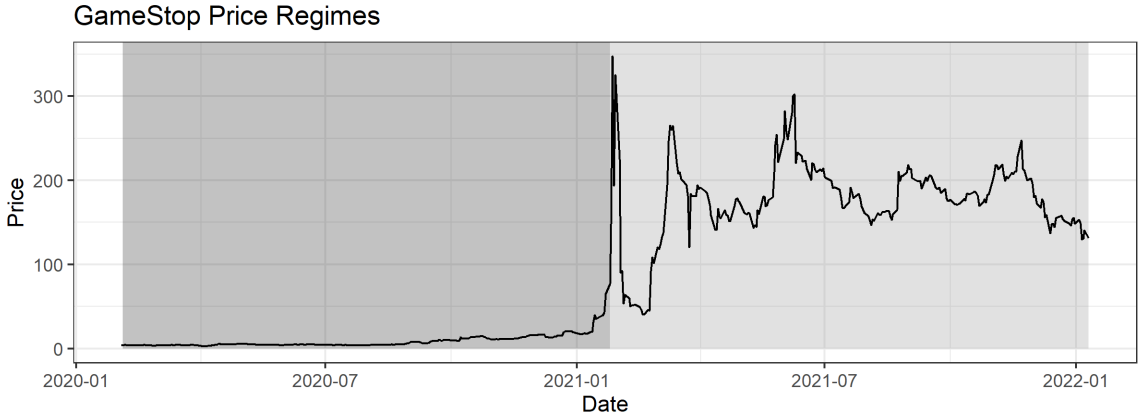

Figure 4.17: Plot of the regimes for GameStop closing price. Dark grey represents regime 1, light grey represents regime 2.

50% of the time, considering there are two regimes possible.

Tesla

Let's look at a stock now where we have put a prior on it not being a meme stock: Tesla. The results for Tesla are the following for the $ln(\#$ posts) time series.

```
Coefficients:
        (Intercept)(S)    Std(S)
Model 1       1.859721 0.6555975
Model 2       3.480833 0.6156020

Transition probabilities:
           Regime 1   Regime 2
Regime 1 0.97238804 0.05270496
Regime 2 0.02761196 0.94729504
```

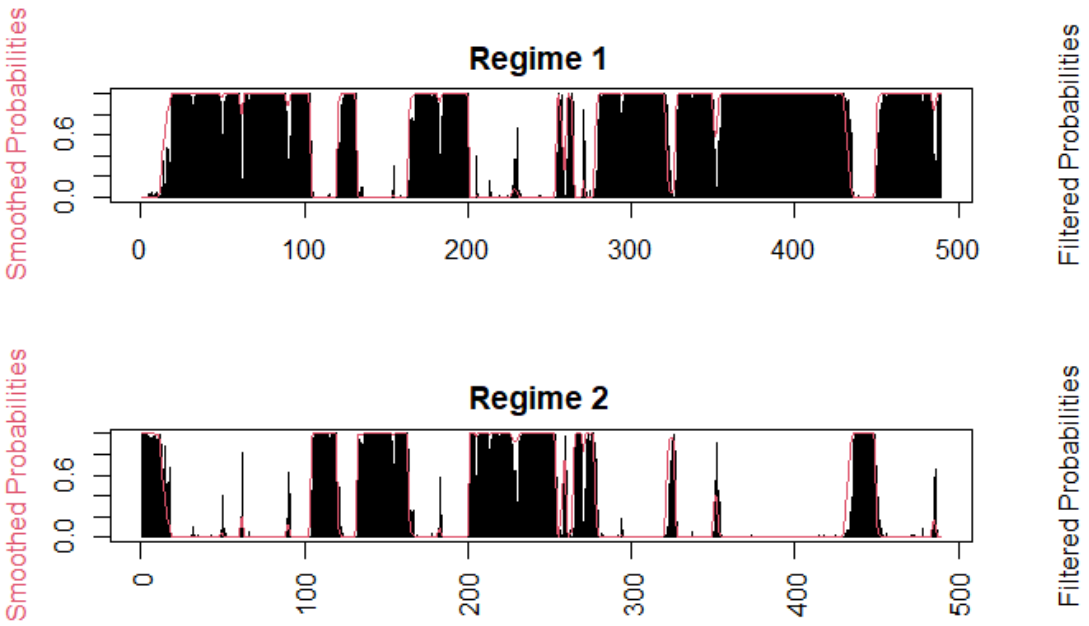

Figure 4.18: Probability of being in regime 1 or regime 2 at each point in the Tesla $ln(\#$ posts) time series.

Here (4.19), the main difference between the regimes are the means. Regime 2 has about double the mean of regime 1, yet both regimes are very close in volatility, with regime 1 slightly edging out regime 2.

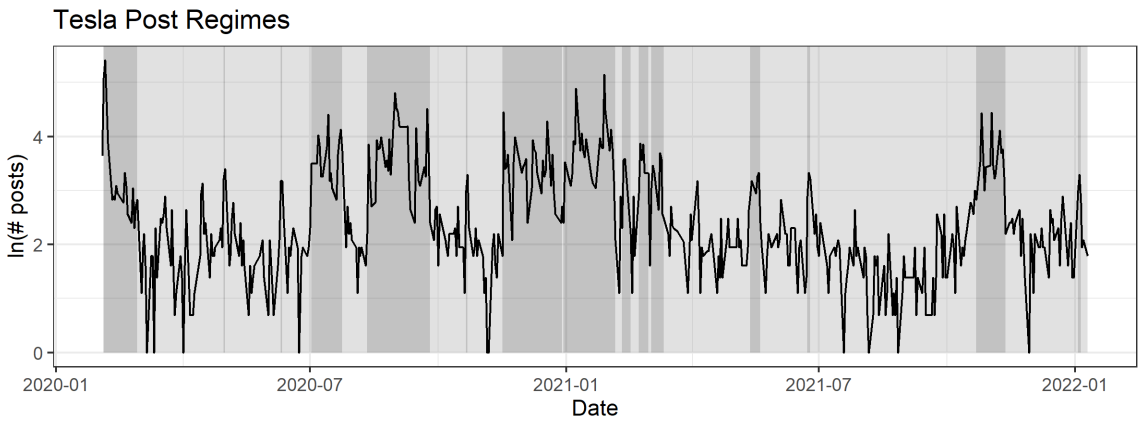

Figure 4.19: Plot of the regimes for GameStop closing price for the entire time period. Light grey represents regime 1, and dark grey represents regime 2.

Tesla's post regimes look quite sporadic compared to what we saw in GameStop. This make sense because there was never an obvious shift in the number of Tesla posts on Reddit like there was for GameStop. Moreover, the volatility shift was essentially negligible between the regimes here whereas the volatility change in the GameStop regimes was more than 2x.

Next, let's look at the regimes for the Tesla price.

```
Coefficients:
        (Intercept)(S)   Std(S)
Model 1       770.6503 166.6022
Model 2       263.9653 125.7095

Transition probabilities:
           Regime 1     Regime 2
Regime 1 0.99650706 2.191144e-13
Regime 2 0.00349294 1.000000e+00
```

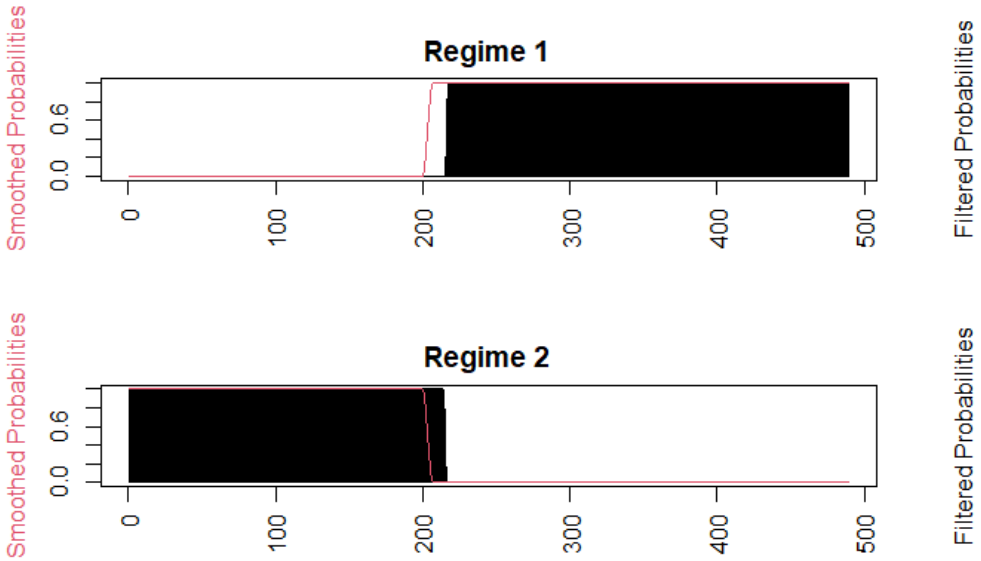

Figure 4.20: Probability of being in regime 1 or regime 2 at each point in the Tesla price time series.

Tesla closing price has just one regime change in totality. Throughout this time period, Tesla steadily rose, which lead to only one estimated regime change. When looking at most price series, the regime with the greater volatility will also be the regime with the higher mean. For Tesla, this was not the case. Instead, the regime with the greater mean was also the regime with the smaller volatility.

The greater volatility regimes in both the price time series and the reddit post time series occurred on the same days only 57% of the days in the series, and the

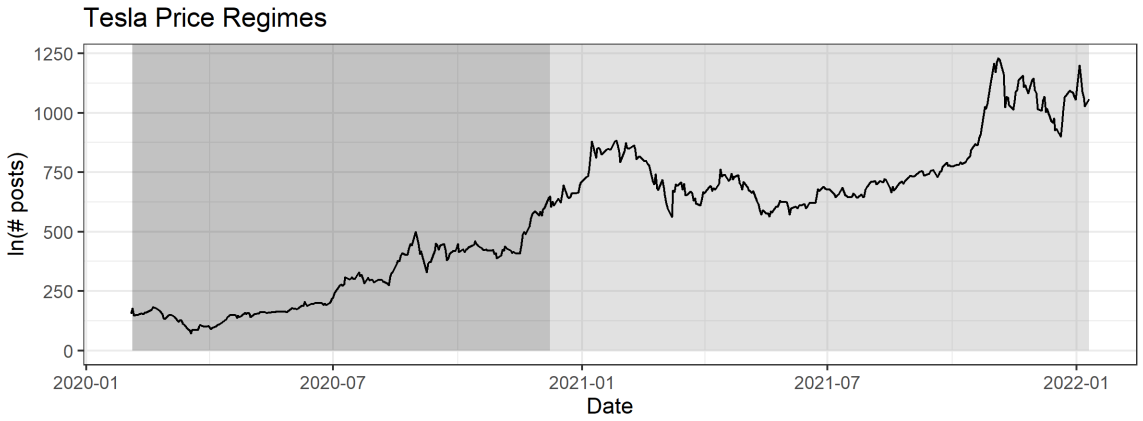

Figure 4.21: Plot of the regimes for Tesla closing price for the entire time period. Light grey represents regime 1, and dark grey represents regime 2.

greater mean regimes in both time series only occurred on the same day 43% of the time. This is evidence that there is not a strong link between the number of TSLA posts on reddit and the Tesla stock price, which aligns with prior expectations.

4.6.3 Compiled Results

The compiled results for the entire list of stocks can be seen in (4.22). The results for GameStop and Tesla have already been discussed, but can be seen in Figure 4.22 as well.

Like stated previously, if each day in each series uniformly randomly chose which regime to assign itself to, then matching regimes in each of the two series would result in a close to a 50% accuracy. Thus, we should expect for the stocks with no relation to the number of reddit posts for that stock to have price regime and reddit regime alignment 50% of the time, because the chance for the regime which the price exists in to be the same regime which the post data exists in is a 50% chance.

Unsurprisingly, the highest and second highest regime alignments come from GameStop and AMC respectively. For GameStop, the regime which had a higher volatility for both price and reddit data also was the regime with the greater mean,

Percent of Days With Same Regime in Price/Reddit		
	Regime Alignment	
Ticker	Mean	Volatility
AAPL	47.0%	47.0%
AMC	81.6%	81.6%
AMZN	54.4%	45.6%
CVNA	44.6%	55.4%
GME	87.7%	87.7%
KO	39.7%	60.3%
MRNA	55.2%	55.2%
NOK	58.5%	41.5%
NVDA	60.9%	60.9%
PLTR	56.6%	43.4%
SAVA	78.3%	78.3%
SPCE	65.6%	65.6%
TSLA	42.9%	57.1%

Figure 4.22: Each percent represents the fraction of days where the regimes line up in the price series and the post series for each stock. Mean regime alignment means that the regime with the highest mean in both series is matched up. Volatility regime alignment means that the regime with the highest volatility in both series is matched up. In most cases, the regime with the highest volatility will also be the regime with the highest mean in both series, meaning that the percents will be the same for each regime alignment. The stocks highlighted in blue are the pre-labeled meme stocks.

making its accuracy (87.6%) in both categories. Surprisingly, several pre-identified meme stocks fell short in this analysis. Both Palantir ($PLTR) and Nokia ($NOK) had no significant regime alignment between their prices series and reddit post series.

Both stocks had less volatility later in their price series, which lead to regime 1 having the greater volatility, and regime 2 having the greater mean.

Every pre-identified non meme stock hovered around the 50% area as expected. Nvidia ($NVDA) was the best performer among the non-meme stocks with a 60.9% alignment in both the max mean and max volatility regimes, but still nothing very high.

Volume

Now, we will discuss the results for volume in lieu of price. The procedure was the exact same as with the price time series and the ln(number of posts) time series like we just saw, but now we will instead use the volume data instead of the price data.

In contrast with the price data, the regimes with the higher mean were almost always also the regimes with the higher volatility as seen by the identical mean and volatility numbers in (4.23). This was in fact the case for every stock on the list except for Tesla. Similar to the regression results, AMC seems to be the most aligned with volume traded and reddit activity, much more so than GameStop. The high volatility/mean regimes for volume and reddit posts aligned on 91.2% of the days for AMC. This is the top accuracy for any stock.

The other notably high accuracy comes from Virgin Galactic ($SPCE) which now has an 82.2% regime link between volume and reddit posts, compared to its 65.6% regime link between price and reddit posts.

GameStop, like in the regression, took a substantial dip in the accuracy of volume and reddit compared to price and reddit. GameStop's volume regimes only align with the reddit regimes 60% of the time, a number in the range of many non meme stocks. Apple, Carvana, Moderna, and Nvidia all amazingly topped GameStop in this category. Among those stocks, Moderna leads the pack with a 66.5% regime link.

Percent of Days With Same Regime in Volume/Reddit		
	Regime Alignment	
Ticker	Mean	Volatility
AAPL	61.1%	61.1%
AMC	91.2%	91.2%
AMZN	53.2%	53.2%
CVNA	64.2%	64.2%
GME	60.3%	60.3%
KO	56.4%	56.4%
MRNA	66.5%	66.5%
NOK	70.1%	70.1%
NVDA	60.5%	60.5%
PLTR	42.9%	42.9%
SAVA	68.9%	68.9%
SPCE	82.2%	82.2%
TSLA	68.1%	31.9%

Figure 4.23: Each percent represents the fraction of days where the regimes line up in the volume series and the ln(number of posts) series for each stock. Mean regime alignment means that the regime with the highest mean in both series is matched up. Volatility regime alignment means that the regime with the highest volatility in both series is matched up. In most cases, the regime with the highest volatility will also be the regime with the highest mean in both series, meaning that the percents will be the same for each regime alignment. The stocks highlighted in blue are the pre-labeled meme stocks.

Chapter 5

Conclusion

The meme stock surge was a phenomenon that we have never seen before. With the power of social media and the internet, millions of small retail traders were able to move massive volumes and stock and send the share prices of these stocks to places many deemed impossible. In my analysis, I have presented evidence that the r/wallstreetbets subreddit had a substantial connection to the price and volume of several stocks. Regular regression results give nearly a $R^2 = 0.6$ when regressing GameStop share price on the reddit predictors. Moreover, the OLS results where volume is regressed against the reddit predictors yield around a $R^2 = 0.8$ for AMC, while finding non meme stocks Amazon, Apple, Nvidia, Carvana, and Coca-Cola to have $R^2 < 0.1$.

Regime analysis provided similar results as the regression, resulting in 91.2% regime alignment with AMC's reddit data and volume traded, and 87.7% regime alignment with GameStop's reddit data and share closing price.

There are several relevant questions regarding this topic that remain yet to be answered. Firstly, who exactly profited from all of this rapid price movement? Sure, the retail traders fueled the initial surge, but the details beyond that are unclear. It's entirely possible that the retail trading population as a whole took a big hit if they

bought too high and refused to sell as the price inevitably decreased. This, in turn, would have allocated the bulk of the profits to the high frequency traders or other such institutional traders who managed to capitalize on the events.

Moreover, are there going to be more surges like this in the future with new stocks? Nothing like the initial GameStop surge has occurred since, but the price seems to still be rejuvenating itself anytime it dips too low still today, indicating that this might be a never ending process as opposed to simple surges.

Appendix A

Code

This is the method I used to query historical reddit data from Pushshift.

```
def getPushshiftData(q, after, before, subreddit):
# Keyword to query --> q
# in between dates ‘before’ and ‘after’
# in the subreddit ‘subreddit’
url = ’https://api.pushshift.io/reddit/search/submission/?title=’+\
str(q)+’&size=1000&after=’+str(after)+\
’&before=’+str(before)+’&subreddit=’+str(subreddit)
r = requests.get(url)
data = json.loads(r.text)
return data[’data’]
```

Bibliography

[1] https://www.wsj.com/articles/judge-dismisses-meme-stock-lawsuit-against-robinhood-and-citadel-securities-11637265778.

[2] https://pypi.org/project/yfinance/.

[3] https://files.pushshift.io.

[4] A. Aloosh, H. Choi, and S. Ouzan. On the efficiency of meme stocks. *Capital Markets: Market Efficiency eJournal*, 2021.

[5] M. Costola, I. Matteo, and S. Carlo. On the "mementum" of meme stocks. *Economics Letters*, 207:110021, 2021.

[6] D. A. Dickey and W. A. Fuller. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, pages 427–431, 1979.

[7] R. F. Engle and C. Granger. Co-integration and error correction: Representation, estimation, and testing. *Econometrica*, 55:251–276, March 1987.

[8] S. M. Goldfeld and R. E. Quandt. A markov model for switching regressions. *Journal of Econometrics*, 1973.

[9] C. Granger and P. Newbold. Spurious regressions in econometrics. *Journal of Econometrics 2*, 1974.

[10] J. D. Hamilton. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, March 1989.

[11] J. D. Hamilton. *Time Series Analysis*. Princeton University Press, January 1994.

[12] M. P. Murray. A drunk and her dog: An illistration of cointegration and error correction. *The American Statistician*.

[13] K. I. a. Park. *Fundamentals of Probability and Stochastic Processes with Applications to Communications*. Springer International Publishing :, Cham :, 2018.

[14] P. Phillips. Understanding spurious regressions in econometrics. July 1985.

[15] J. A. Sanchez-Espigares and A. Lopez-Moreno. Mswm examples. June 2021.

[16] E. Sims. Notes on time series. 2013.

[17] L. P. H. Yacine Ait-Sahalia. *Handbook of Financial Econometrics*. Elsevier, 2 edition, 2010.