2019

Market Manipulation and a Case for the Further Regulation of Social Media and the Finance Industry

Ross Powell
Skidmore College, rpowell@skidmore.edu

Market manipulation and a case for the further regulation of social media and the finance industry

Ross M. Powell

May 1, 2019

Abstract

This paper intends to investigate the potential of market manipulation in under regulated markets that does not exist in regulated ones. I do this by looking at the previous literature discovered on ambiguity aversion, and how it is linked to the increase of social media's effect on price changes in markets. I investigate two different markets that are regulated and unregulated. These markets are the US stock exchange and the cryptocurrency market. I then see what the effect that twitter has on the two over the same periods of time using the most up-to-date models. Finally, I recommend policy changes that will help prevent market manipulation of under regulated markets.

Introduction

A discussion of media and market manipulation

Some of the biggest problems we have today stem from poor regulation of media and industry. The 2016 election is a great example of how poor media regulation could have led to manipulated outcomes. One of the main discoveries was that media influence specifically on social media sites could be harnessed to promote false narratives. Multiple studies confirm the populous' vulnerability to social media manipulation. One such study, Gottfried and Shearer (2016), found that 62% of adults get their news from social media. This means that there are more people who get their news from social media than people who get their news from centralized news agencies. In terms of examples where media manipulation has been used we need to look no further than the 2016 election. According to Silverman (2016), the two most popular "fake news" stories were more widely shared on Facebook than the most popular mainstream news stories. This means that outside agents placing fake accounts on social media sites to promote false narratives had more views than the most popular mainstream news stories at the time. In tandem, these two discoveries paint the picture that there are almost two different worlds between media and social media. In the case of the 2016 election each of these "worlds" had an equal power in voting, which influenced the elections in ways that we were unable to predict.

Market manipulation is a whole other side of the power of disinformation. A specific case of this is found in Gandal et al. (2018) with the market for Bitcoins. According to their findings there were several repeated buys that caused media excitement and a buy frenzy which drove the price of bitcoin six fold. The manipulation here was in the form of a pump and dump scheme, but the media excitement that followed the buy orders clearly shows the effect that media has on the general public. This phenomenon is less of a problem when a market has centralized news sources, such as the television and radio stations of MSNBC, FOX, and CNN. The reason for this is that the company's credibility and ability to succeed are primarily driven by accurately telling the news. This is because credible mainstream television stations are more likely to succeed when they tell accurate news. Proof for this is shown in constantly reoccurring polls that look for consumer confidence in the top television stations, where most television stations that rank highly in confidence are correlated with higher rates of viewership.

Industry regulation is the other issue at hand. If the industries that media can manipulate are already heavily manipulated in the first place, the regulation of media won't really make much of a difference. For instance, think of the finance industry during the Great Recession in 2008. The Great Recession was caused largely by a group of people that dealt with the packaging of collateral debt obligations or CDO's. Purchasing a collateralized debt obligation means that an investor buys up a part of the debt owed by a homeowner, with the idea that when the homeowner pays the debt back, the investor gets interest and makes money. Normally this process works if the ratings agencies are accurately classifying the mortgages used in a CDO, but many of these agencies, in order to stay in business, had to rate all the mortgages highly or else a bank would just go to a different ratings agency. This shows that the banks had far too much bargaining power with the ratings agencies. As this process continued the media caught onto the mortgage hype and propelled it further which created a cycle. Eventually, it all fell apart. Homeowners began defaulting on their loans because they did not have good credit.

A key takeaway here is that any reliable media source would see that most CDO's were made on "good" loans, given that a vast majority of the trusted agencies had nothing but favorable information about the loans. The media would continue to report that the housing market is safe, and a great way to make money which influences consumers. At any rate, the media was not blame for the financial crisis in 2008. To prevent structural problems like this, the regulation of industry is crucial in the development of fair markets. Therefore this paper will focus on the regulation of industry in addition to social media, which was outlined above.

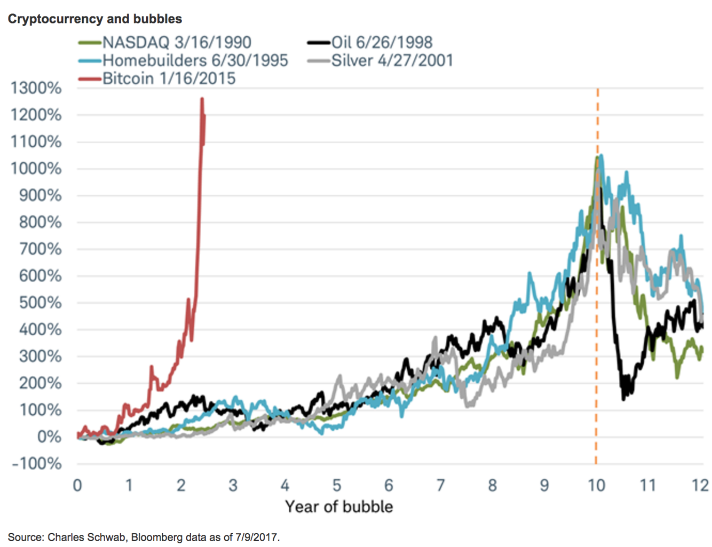
What will this paper focus on?

The specific industry that I am going to focus on is the regulation of the finance industry. This form of regulation is of utmost importance because the finance industry is vital to the success of the current world economy system. The US stock market is an example of a regulated subsection of the finance industry that is generally free of corruption due to organizations like the Securities and Exchange Commission (SEC). With little fraud, the US stock market has grown immensely, allowing for many people and businesses around the world to make money together. On the other hand, markets with little regulation are prime targets for manipulation. Markets like these can lead to highly volatile markets where often a select few people have an enormous amount of market power. An example of a market like this is the cryptocurrency market.

Before I start explaining the nuances of the cryptocurrency market, I'd like to start by introducing the basics of what cryptocurrency is. Cryptocurrencies propose a unique approach toward financial infrastructure. Technological solutions like secure peer-to-peer connectivity and decentralized ownership allow for a different approach than our current centralized monetary system. Cryptocurrencies first arrived on the world's doorstep via Nakamoto (2008) in the wake of the Great Recession of 2008, mostly in response to numerous government bailouts and rumored stimulus spending. They promised to provide people safety from federal reserve polices and the devaluation of the dollar, an attractive option to some after witnessing the current banking system fail to prevent fraud. The main cryptocurrency in the early years was Bitcoin. It was designed to encrypt money transfers, and to provide a decentralized version of the financial system so that no governing body had the ability to manipulate the supply of currency.

From its inception to its peak, Bitcoin's value increased more than one thousand fold. Its price increase was explained by it's limited supply, and that it could potentially replace the whole banking system. As the news spread, owners started to treat Bitcoin and other cryptocurrencies as a major speculative investments instead of their intended use: mediums for exchange. Phrases like "HODL" (a mistype of the word "hold") and "To the moon!" were seen all over online message boards acting as signals to others that the world believed in a never ending increase in cryptocurrency value. All these reasons, in tandem with a young population of internet users with access to cryptocurrency exchanges lead to one of the biggest market bubbles ever recorded as shown in Figure 1.

Figure 1



The cryptocurrency craze happened for a number of reasons. The main reason was organized trading networks in conjunction with seemingly limitless trading created an environment perfect for pump and dump schemes. A majority of the rapid increases were caused by cryptocurrency owners that had very large percentages of the total circulating coins called "whales." Often times a whale would set up a bot that would buy up high amounts of coins, driving the price up. The rise would be reported on by media and would excite many people that would then buy the currency too. After the whale would sell the currency back to the unsuspecting people and the price would collapse, even though the whale attains more money and cryptocurrency. An example of this is in Ganal et al. (2018), where the authors show that it is very likely that a single actor caused the bitcoin price to jump from \$150 to \$1000 in late 2013. Continued hype over these sorts of rises caused the general public to buy cryptocurrencies at irrationally high prices, and in extreme cases take out a second mortgage to buy it.

There are many ways a market can get over hyped. Since the cryptocurrency market is mainly reported on by twitter and various cryptocurrency sites, the opinions of many enthusiasts outweighed the main stream news. Many of these sources do offer the news on cryptocurrencies along with the most up-to-date prices, but often times write very polarizing opinions on cryptocurrency. The problem was that these were pretty much the only opinions users could get during the bubble, so it created a feedback loop causing extreme market volatility. On the contrary we can compare the mainstream media. Yes, they get it wrong, and sometimes really wrong like their reporting leading up to the Great Recession, but overall they are reporting the news given the information that they have, which is what they did leading up to the Great Recession.

Another key issue hinted at above, is who is trading the assets within a market, and who owns the assets within a market. According to Brock et al. (1992), forty-five percent of US stock market trades made in 1990 were done by institutions – a number that is likely higher due to the rampant growth of the financial industry over last couple decades. However, it is important to note that the money being traded by institutions then and now is not usually not owned by them, and the institutions usually have to go through a lot risk mitigation in order to ensure that they are not making bad bets with other people's money. The cryptocurrency market is quite different. Roughly 40% of the market is owned by an estimated 1000 people according to a study done by AQR Capital Management [1]. When this is the case these people can price manipulate causing uneducated people to panic and buy in or sell off when the true nature of the market would not indicate doing so.

[1] The author Aaron Brown is a contributor on the Bloomberg prophet column. I found it sited at https://www.bloomberg.com/news/articles/2017-12-08/the-bitcoin-whales-1-000-people-who-own-40-percent-of-the-market

In tandem, these points paint the picture that no markets are perfect, but under regulated ones like cryptocurrency are likely to be targeted for manipulation. In order to narrow down my study on industry and media manipulation, I will focus on the US stock market and cryptocurrency market. This permits me to investigate the differences in regulated and under regulated markets. I will focus on the use of media as a potential tool for price manipulation. I will try to make an argument that we should identify some of the major problems with social media and suggest regulation. To show this empirically I will use models used in the economics literature for indicating price shifts in markets, and compare the results. My research hypothesis is that social media accounts for more variance in the cryptocurrency markets than regular markets, and further regulation of social media and the underlying markets are a necessary precaution in preventing market bubbles.

Literature Review

I am now going to connect this topic to the behavioral economics literature. I'd like to start by honing in on how the news and media effect our opinions. The main explanatory theory here is ambiguity aversion, which put simply is an agent's preference for certain decisions over uncertain decisions. After discussing ambiguity aversion and where in the literature it comes from, I will go over how investment outcomes are usually ambiguous and how media is sought after for informing investment decisions. Finally I will go over models used in the literature currently for modeling markets with social media in order to get support for my analysis.

The history of ambiguity aversion and how it ties into media's influence

Ambiguity aversion's history is rich and it's something that effects us all the time. The origins of this theory stem from expected utility theory which was first discussed in the eighteenth century by Daniel Bernoulli. The issue was talked about over the years but it wasn't until the early twentieth century that this theory started to become more mainstream. One of the more famous papers of this time that discussed expected utility theory was Neumann and Morgenstern (1944). They claimed that expected utility theory is a normative theory of behavior, and that it fit into the classical utility theory of behavior. Classical utility theory assumes that agents are rational and utility optimizing. Therefore the expected utility is the amount of utility an agent will likely receive from an event, while also knowing the likelihood of the event happening. Expected utility can be modeled by the following example. Assume that an agent gets to choose two different places to travel, location A and location B. On location A there is a 50% probability of event C or D happening. On location B there is 70% chance that event C happens, and a 30% chance that event D happens. The agent is rational and wants to optimize his utility. The utility the agent derives from event C happening in either location is 2, and the utility the agent derives from event D happening in either location is 5. Solving for the expected utility of choosing location A and B

$$EU = P(C_A) * U_c + P(D_A) * U_d = 0.5 * 2 + 0.5 * 5 = 3.5$$

$$EU = P(C_B) * U_c + P(D_B) * U_d = 0.7 * 2 + 0.3 * 5 = 2.9$$

A rational agent would therefore choose location A, because A's expected utility is higher than B's. This model has all sorts of applications when the probabilities of an event happening are known, and it is a great starting framework for understanding classical decision making with probability. One limitation of expected utility theory is that not all economic agents are rational, and that expected utility theory does not have a way to measure uncertain probabilities. In many case probability distributions of an event are rarely fully known to us when we make decisions. This means we have to have a way to estimate probability distributions of events without actually knowing any of the probabilities. The name for this sort of assessment is called the "subjective probability assessment." It is essentially our mind coming up with the best answer it can given what it already knows. This idea is first modeled in Savage (1964) where he develops a counterpart to expected utility known as subjective expected utility (SEU). SEU can be thought of as an individual's expected utility function, weighted by his or her subjective probability assessment instead of a discrete probability. An equation for SEU is as follows:

$$SEU = SPA(C_A) * U_c + SPA(D_A) * U_d$$

This equation is similar to the equations used above modeling expected utility, but instead of discrete probabilities being known, which I denote as P, we have subjective probability assessments, which I denote as SPA. In words this equation resembles an agent's subjective probability assessment of events C and D happening on location A. It's key to note that the probabilities are uncertain, and that the user has guessed them given the information present.

One of the first papers contrasting expected utility theory and subjective expected utility theory was Ellsberg (1961). Ellsberg conducted an experimental model that showed the general population is averse to ambiguous probability distributions. In other words the general population of the experiment model preferred expected utility theory over subjective expected utility theory. In this experiment participants were told to choose an urn they would like to select a ball from. Each of these urns contained 100 balls, where Urn 1 had 50 red balls and 50 blue balls, and Urn 2 had a random distribution of red and blue balls. The participants were first given the choice of trying to find a red or blue ball, and then were given the choice of the urn that they would like to select from. They were also told that they would win $100 if the color of the ball selected matched the color of the ball they were trying to find. In both urns for both colors the Bayesian probability is the same (50%). However, 84% of the participants choose Urn 1, showing a general aversion to the ambiguous choice.

Another example that accurately shows the curiousness of this phenomenon is shown in Barberis and Thaler (2003). In this paper the authors used an experimental approach to show ambiguity aversion. The author's made a game where a participant was asked to give a subjective probability assessment of some random event happening. After the participant gave their guess, they were told that a machine will take that probability and make a game itself. In this game the machine has a discrete probability equal to the original guessed probability of flashing 1, and flashed 0 for an unsuccessful guess. The participant was then given two choices, the original game where the probability of the random event happening was unknown, and the machines game where the probability is certain. In almost all cases the participant chose the machine's game, likely due to the probabilities being discrete and not ambiguous.

The final paper I would like to have in my discussion of ambiguity aversion is Maccheroni et al. (2006). Maccheroni et al. cite Hansen & Sargent (2000) in their talk about how ambiguity aversion plays a role in our most basic models of uncertainty. Hansen & Sargent make the case that bad information received by agents being ambiguity averse is what informs agent's opinions on the base of their model for bad decisions they make under uncertainty. This link was hypothesized mostly in Hansen & Sargent (2000) but is subsequently explored further within Maccheroni et al. (2006). Here the authors give solid grounds for the link connecting ambiguity aversion to choosing bad information, which is what Hansen & Sargent (2000) set out to uncover in the first place. Furthermore Maccheroni et al. (2006) developed an analytical framework for ambiguity aversion where the utility function is weighted by an ambiguity index on the set of probabilities of events. The result was a utility function that captured the agent's risk preferences, weighted by an ambiguity index that captures the ambiguity attitudes that an agent might have. I won't use this sort of model in my analysis, but this does show how literature has established the link between ambiguity aversion and bad information.

The link here is that agents may take in bad information when making investment decisions that they are uncertain about. In order to make this probability distribution less ambiguous agents will do research on their market by gathering opinions and facts to inform our own. This means that when we are uncertain of outcomes our opinions are largely derived from where we receive our information from. For instance think of the 2016 election, and an average uneducated citizen that wants to vote. This citizen wants to conduct research of their own, and stumbles upon disinformation that informs their opinion. They then vote in the election due to disinformation. Similar links can be made between under educated investor and investment decisions. In general most investors don't have the time or capability to make complex investment models, so they research information to inform their opinions. Information that they stumble upon may be disinformation and can also lead to faulty investment decisions. In recent years this issue has been brought more to the limelight. The cryptocurrency bubble and the 2016 election sparked a lot of interest in specifically social media's ability to influence important world events.

Other contributing factors to social media's ability to manipulate markets

One of the biggest contributing factors to social media's ability to manipulate markets is caused by human nature. Social media sites provide an environment where a user gets to select what news and which opinions show up. Naturally we select the things that we agree with and that confirm our previously held beliefs. In addition to selecting the news, social media allows us to select our friends. This means that people we disagree with are often times underrepresented in our feeds, which further confirms our previously held beliefs. Del Vicario et al. (2016) found evidence for these two claims and show that social media sites always have groups of people that aggregate around common interests and beliefs. As the authors point out, this functionality is not bad and is the whole point of social media, but it creates an environment where users start to trust each other within interest groups. When Del Vicario et al. (2016) looked into the sources of misinformation, they founds that misinformation is often times spread by a single user within one of these groups, which then cascades out into other groups. This is interesting because it promotes the narrative that social media groups are more at risk for misinformation than other forms of media. Additionally Mocanu et al. (2015), found evidence that continued exposure to misinformation caused users to believe higher amounts of misinformation. This means that groups that have a source of misinformation will overtime be less aware of the likelihood of a story being misinformation. This is important to note in my analysis because cryptocurrency enthusiast groups might have been exceedingly more at risk for misinformation as the bubble grew. In tandem these papers show that social media users are more vulnerable to misinformation than other mainstream media sources due to structural reasons in how users organize themselves.

Analytical frameworks that model these behaviors

As shown above the literature has plenty of evidence to support that ambiguity aversion is linked to an agent's choice to seek out bad information. I will deduce this to saying that ambiguity aversion causes agents to seek out information that impacts their decision making process. Furthermore, I've shown that social media has the optimal structure for misinformation due to groups forming around common interests. I will deduce this to saying that social media has the optimal structure for spreading potentially polarizing views. Together these two points show how ambiguity aversion leads users to seek out potentially polarizing information in order to inform their own views. In order to analyze these effects I researched papers that used social media signals to predict market price movement. One such study that sought to analyze this was Nguyen et al. (2015). Their goal was to create a model which predicted stock price movement from message boards on Yahoo! Finance. Their main contribution was that they proposed a 'topic-sentiment' to improve their model's prediction capabilities. They use two datasets – one on historical stock prices, and another that displayed the content from the message boards. Their algorithm was able to accurately capture roughly 54.4% of direction of price movement. However, some stocks in their sample were predicted more accurately, such as Amazon (AMZN) for 71.05%. The main limitations of this study is that only 18 stocks were picked for the sample, and their data scraping method for finding "topics" to portray sentiment was (by the authors' own admission) fairly weak, as there were a limited amount in the study.

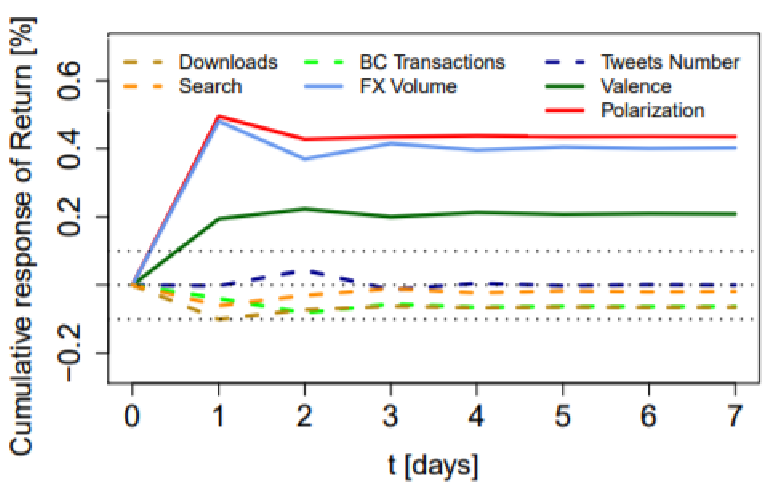
Unlike Nguyen et al. (2015), who use message board data to predict stock prices, Ranco et al. (2015) use Twitter data to measure the same thing. The authors collected data from the 30 stocks from the DJIA over a 15-month period. For each of these companies, the authors built a time series of general sentiment in financial tweets. They restricted their analysis period on stock returns to shorter periods around certain "events" like releases of earnings reports. The main findings of this study show that there is significant evidence that Twitter sentiment is a driving force behind stock returns around event periods. This paper is one of many examples that show that social media sentiment can influence market behavior.

For Bitcoin specifically, Garcia & Schweitzer (2015) examine its price fluctuations in conjunction with social signals such as social media action and online polarization. They develop a framework to detect price changes via an algorithm that develops criteria on which signals occur before market fluctuations. This framework paves the way for the authors propose a systematic trading strategy to exploit these behaviors in the Bitcoin ecosystem. The biggest social signal that influenced price changes in this study was the volume and sentiment of tweets related to Bitcoin, showing that twitter sentiment plays a large part in Bitcoin's price changes, again providing solid grounds for my study.

Important variables given by the literature

The model used in Garcia & Schweitzer (2015) is a good starting point for modeling twitter sentiment. Their model included several measures for twitter sentiment and key market variables. The main measures they indicate as important for twitter sentiment are the number of tweets and general attitude. The authors got the number of tweets by searching the number of times the word "Bitcoin" was in a tweet over a day. The next variables are the twitter polarization and valance. The twitter polarization is what we think of when we say sentiment. It is essentially a score of how many tweets are positive to how many tweets are negative. If the there are a lot more optimistic tweets, the polarization is considered positive. If there are a lot more pessimistic tweets this obviously means the polarization is negative. The valance is how emotional the tweets are. If the tweets are considered more emotionally driven, there's likely to be more volatility. The results for how these three effect the return variable (used in Garcia & Schwietzer 2015) are shown in Figure 2.

Figure 2



As mentioned before this paper also includes more general market variables. These variables include the price change of Bitcoin, the exchange volume, and for bitcoin specifically, the total amount of bitcoin transactions. The price change of Bitcoin can be attained by subtracting the difference in prices over a specified time period. The exchange volume is essentially the amount of bitcoin that was traded over a 24 hour period. The final variable is the number of transactions which obviously tells us how many trades happened over the course of a day.

Using the work from Ranco et al. (2015) in tandem with Garcia & Schweitzer we get additional support that the overall attitudes are important in the US stock market as well. Connecting this back to the literature we see that we can use these models to test for the prevalence of twitter in determining the price change, and then to what regulatory policy changes might be necessary in order to try and prevent fraud in unregulated markets like cryptocurrency.

Data and Methodology

I will be utilizing twitter data, price change data from the DJIA as well as price change data of the cryptocurrency market. Following this I will clean my data so that I can conduct two separate regressions: one for the cryptocurrency market, and one for the DJIA. By comparing the results we will see the effect of social media on each of these markets.

The process to extract this data is laborious since it all came from different places. The tweets came from the Twitter API, which can be thought of as a massive twitter app where each tweet is stored. The Twitter API offers a several levels of access. The first is the free method which allows users to stream as much data as they want, that is to capture and store what is being tweeted in real time. The other options all include large payments in order to use twitter's historical database. The latter is the most useful in my analysis, but given it's cost I could not use it. Instead I streamed tweets, which is another way to do this project. In order access the free version, I had to get an API access token. Upon getting an access token, I started searching for programming languages that had dense twitter libraries. I found that the R language's twitteR[2] and rtweet[3] packages seemed to be the most useful because they have robust functions that allow for keyword streaming.

The keyword streaming part of this project was tough, and through trial and error I found the way to most effectively capture data that was useful. The keywords "cryptocurrency" and "crypto" were the two main keywords that I used. They are some of the most likely keywords to find cryptocurrency related tweets, and also some of the keywords that were the most unlikely to be mistaken with anything else. I also used the keyword "hodl," a common phrase used around the cryptocurrency space, and the names of the top three cryptocurrencies: Bitcoin, Ethereum, and Ripple. I used specific currencies here because they comprise of roughly 60% of the cryptocurrency market cap, which can be used as a proxy term for the cryptocurrency market. Overall these keywords work well and will provide accurate tweets relevant to the cryptocurrency market. The DJIA keywords followed a similar format to the cryptocurrency keywords, so the keyword "dow jones," various forms of the acronym DJIA, and the acronyms AAPL and BA for Apple and Boeing respectively were used. I'd also like to note that I used the same selection process in choosing these keywords for both the cryptocurrency and djia in order to keep the process consistent and less biased.

The process of streaming is trivial. I streamed the data from 2PM to 3PM EST during weekdays. I did this because the twitter sentiment would likely be more reflective on the day's sentiment since the opening bell price movement had happened, and people would have time to react to that. With cryptocurrency this did not matter since the market is always open, but to keep it consistent I collected data for both during this time. Following streaming the data, I had to create a way to decide the overall sentiment of the tweets on a given day. I was able to effectively do this using R's syuzhet[4] package which calculates which emotions are in the tweet, uses matrix multiplication to reduce the result to one

[2]Details here: https://cran.r-project.org/web/packages/twitteR/twitteR.pdf
[3]Details here: https://cran.r-project.org/web/packages/rtweet/rtweet.pdf
[4]Details here: https://cran.r-project.org/web/packages/syuzhet/vignettes/syuzhet-vignette.html

number, and finally judges the tweet to be positive, negative or neutral.

After I got a score for all tweets, I had to break down each day's data to the number of times a tweet showed up as positive, negative and neutral. When I had these numbers, I could calculate the day's sentiment by seeing what proportion of the tweets were positive of the whole. The thresholds I established for daily proportion x are defined below:

$$\text{Classification} = \begin{cases} \text{Positive (1)}, & \text{if } x \geq 0.66 \\ \text{Neutral (0)}, & \text{if } 0.33 < x < 0.66 \\ \text{Negative (-1)}, & \text{if } x \leq 0.33 \end{cases} \tag{1}$$

Concluding my analysis with Twitter, I have derived the feature "tweetnumber" which will look at how many tweets were found during the hour long time period. These two variables are consistent with what was used in Garcia and Schweitzer (2015) as the biggest predictors of social media's effect on price change.

The next step was finding a reliable source for the daily price changes and overall market volume. As stated before, the daily price change will be the dependent variable in my study. The overall market volume is the control variable that I will use which is mentioned in Garcia and Schweitzer (2015). In order to be consistent, I had to find a time that I could check both cryptocurrency and the DJIA index for price change and overall market volume. I decided that the range of 9AM to 4PM was best since that is the time that the stock market is open.

The variables I derived are PC for price change, $TSENT$ for twitter sentiment, $TVOL$ for twitter volume, and MV for market volume. I derived these for both the cryptocurrency and DJIA markets. Next I formulated the following model from the literature:

$$PC_t = \beta_0 + \beta_1 TSENT_t + \beta_2 TVOL_t + \beta_3 MV_t + \epsilon_t$$

This model is a representation of the model used in Garcia and Schweitzer (2015), which includes most of the features that I have derived. The only difference other than the model choice is the method that I collected the data, which is not as robust as Garcia and Schweitzer due to my data collection restraints.

Final Data Table

My final data table consisted of 11 different days over the course of a little over two weeks. The first day of sampling was April 1st 2019, and the last day of sampling was April 15th 2019. Ideally I would have liked to have gotten at least a month's worth of data, but my time constraints did not permit me to continue to get consecutive data. On each of the days I calculated the twitter sentiment score, and recorded the number of tweets that showed up. Overall cryptocurrency had a much larger volume that the DJIA, but since the methods remained consistent of the span of the study, the difference in volume did not impact my study design.

My price change variable for the cryptocurrency market was derived from the most popular website in cryptocurrency price tracking[5]. This website has a cryptocurrency market cap tracker which I used to derive the price change of the cryptocurrency market. Each day I subtracted the market cap at 9AM from the market cap at 4PM. The result was the difference in prices which is what I wanted. The price change variable for the DJIA was much easier to find. I found the daily price change on Yahoo finance[6] and recorded the result for every day in my dataset. Additionally, Yahoo finance records the overall market volume, so I scraped the overall market volume for all of the days in my dataset.

The market volume variable for cryptocurrency was found on another major cryptocurrency price reporting company's[7] website. A limitation here is that the market volume was for the entire 24 hour period and not just the 9AM to 4PM window. This does cause a

[5] The website is https://www.coinmarketcap.com
[6] The link is https://finance.yahoo.com/quote/%5EDJI?p=%5EDJI
[7] The link for this is https://coincheckup.com/global

difference within the DJIA and Cryptocurrency datasets, but it doesn't change how the volume is measured for cryptocurrency.

Results

The results for the robust cryptocurrency regression are in Table 1. The p-values are located in the column labeled p, and the observations and R^2 values are labeled accordingly. Twitter sentiment was found to have the most significant relationship with cryptocurrency price change with a p-value of 0.002 indicating that it was significant at the 1% level. This result confirms what Garcia and Schweitzer (2015) found in their study, and gives further credibility to my research question. Twitter volume had an insignificant relationship with price change with a p-value of 0.293, indicating that the quantity of tweets captures did not have a significant effect on cryptocurrency price change. This result wasn't too surprising considering hype can generated by lows and highs. The market volume variable was significant at the 10% level with a p-value of 0.092 and a coefficient of 0.08. This meant that my model found that market volume had a significant positive effect on price change. The Adjusted R^2 is also high, but since my analysis is a time series the adjusted R^2 is biased higher. Additionally the VIF's were all less than 5, indicating that there was no presence of multicollinearity.

The results for the robust DJIA regression are in Table 2. Twitter sentiment had a p-value of 0.389, indicating that it has an insignificant relationship with the price change of the DJIA. This was surprising to me because the results contradicted what was expected from Ranco et al. (2015). Twitter volume was also insignificant with a p-value of 0.629. This result again wasn't too surprising considering how hype can generated by lows and highs. Finally, the DJIA market volume was also insignificant with a p-value of 0.719. This was also surprising to me at first though it intuitively makes sense. Higher market volume is associated with higher volatility which can be positive or negative. The Adjusted R^2

value was -0.173, indicating that the independent variables had no explanatory power with the dependent variable. It is also important to note that the VIF's were all less than 5, indicating that there was no presence of multicollinearity.

Limitations

All studies have their limitations, and mine is no different. The first main limitation I'd like to note is length the data sampling period. Since I only gathered two weeks worth of data, my results did not have enough observations to make a normal distribution. This means that my results may not come out the same in a study that is over a longer period of time. Studies utilizing this framework or similar frameworks to analyze social media's influence on price change should have a sampling period that is longer than the one that I have. The next limitation I'd like to note is that the cryptocurrency market and the DJIA behave differently between the hours of 9AM and 4PM. Cryptocurrency can be traded 24 hours a day, whereas the Stock Market is only open for 7 hours a day during weekdays. This means that shocks felt over the weekend for cryptocurrency were not getting captured in my data which likely biased the results. The final main limitation of this study is that my tweet filtering scheme was only finding a small amount of DJIA tweets which probably didn't give me enough tweets to correctly determine the true sentiment on twitter about the DJIA.

Conclusions

My results show that my research hypothesis was correct. Under regulated markets like cryptocurrency were more prone to social media influence than established regulated ones like the DJIA. This makes sense because social media provides a much larger proportion of the news for cryptocurrencies than the DJIA. There are two major insights to draw from these results. The first major point is that social media sentiment has a very strong effect on the market cap change for cryptocurrency, and likely for other investments that are in under regulated classes. This means that social media does have the ability to manipulate price in under regulated markets. If social media was taken advantage of in a big way to promote buying or selling cryptocurrency, it could cause more pump and dump schemes, hurting unsuspecting investors.

The second major point is that the proportion of the news that social media delivers to the public continues to rise. As stated before in the introduction, social media accounted for 62% of the news consumed by adults (Gottfried and Shearer 2016), which is a shockingly high number given that most social media platforms were first established a little over a decade ago. In addition, the companies that manage all of these platforms have had all sorts of data security issues including the Cambridge Analytica scandal. This is clearly an area that we have work on in order to prevent market manipulation.

In tandem these points show that social media is becoming immensely powerful in providing the news. Additionally, we know that social media has a lot of vulnerabilities that can be taken advantage of. This means that new regulations need to be put on social media companies in order to make sure that their network isn't being used in market manipulation.

Policy Recommendations

Most of the policy recommendations are closely tied to some previous stories in the news, as well as the results of these regressions. First I would like to go over Gandal et al. again, which was cited earlier in the paper. According to Gandal et al. (2018) there was likely a singular actor which caused the first major jump for bitcoin back in 2015. As discussed in the results section of the paper, several repeated buys that occurred on a schedule caused media excitement and a buy frenzy which drove the price of bitcoin six fold. What if these people had access to bot networks on twitter or the other major cryptocurrency media providers as cryptocurrency became more adopted? It would be price manipulation nightmare. Another problem is the presence of trading bots in general. An article from the wall street journal[8] shows several different types of bots that are still legal and fully functional today. "Harassing bots" are known for executing a sell order that is low, prompting a user to buy a cryptocurrency. However just before they offer to buy it, the bot will cancel their sell order, causing the user to buy the cryptocurrency at a higher price, driving the market price up. More notably is when a bot trading network harasses the market, or when a bot trading network trades large amounts of cryptocurrency all at once causing a pump in dump scheme. Clearly there is more to be done with restricting the amount of power these programs have.

Due to these reasons I believe that a way to prevent market manipulation from occurring requires that media providers be more diligent about providing headline stories that are less polarized. In the case of Twitter and other major social media providers like Facebook it is imperative that search queries are focused on diverse, relevant opinions. If a search query can be overrun by hype and bot networks, there is potential for disinformation to be spread, which can likely lead to price manipulation. A solution to this is having a division within the big media companies that keeps track of emerging markets, and makes sure that the result of someone looking for real information leads to a better understanding of the market itself, not a fake one orchestrated by a bot network simultaneously declaring that it is a good time to buy bitcoin. In the case of opinion websites, it is important that the stories are diverse, and that the message boards are monitored for excessive hype and hostile responses towards conflicting opinions. Such a responsibility should be required of all media providers in this regard.

In the case of the underlying cryptocurrency market, much can be done to stifle hype and bot trading. Since the regulation would have to be on the exchanges instead of the

[8]This article outlines several bot types in it, and shows how the cause price manipulation. https://www.wsj.com/articles/the-bots-manipulating-bitcoins-price-1538481600

technology, we can focus on limiting the amount of transactions that a user can make on an exchange back to dollars and other main stream fiat currencies, as well as the purchase of USDT (USD Tether), a cryptocurrency that has the purpose of staying at the value of $1 for purely trading purposes. Together these regulations target bulk traders, and traders that use bots to consistently trade in manipulative ways, instead of real adopters and investors in the technology. Of course the regulations would have to be adopted by all countries, but if they were a lot of the price manipulation in the cryptocurrency market can be helped. Further research will need to be done to show that these policy recommendations would be useful in reducing market manipulation in under regulated markets like cryptocurrency.

Table 1 - Cryptocurrency regression results

	Price Change		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	-3.33	-7.81 – 1.15	0.188
Twitter sentiment	3.90	2.24 – 5.56	**0.002**
Twitter Volume	-0.00	-0.00 – 0.00	0.293
Market Volume	0.08	-0.00 – 0.15	0.092
Observations	11		
R^2 / adjusted R^2	0.791 / 0.702		

Table 2 - DJIA regression results

	Price Change		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	153.92	-679.67 – 987.51	0.728
Twitter sentiment	63.41	-72.02 – 198.84	0.389
Twitter Volume	0.02	-0.94 – 0.97	0.975
Market Volume	-0.70	-4.39 – 2.98	0.719
Observations	11		
R^2 / adjusted R^2	0.179 / -0.173		

References

Barberis, Nicholas, and Richard Thaler. (2003), "A survey of behavioral finance." *Handbook of the Economics of Finance* 1: 1053-1128.

Brock, William, Josef Lakonishok, and Blake LeBaron. "Simple technical trading rules and the stochastic properties of stock returns." *The Journal of finance* 47, no. 5 (1992): 1731-1764.

Del Vicario, Michela, Alessandro Bessi, Fabiana Zollo, Fabio Petroni, Antonio Scala, Guido Caldarelli, H. Eugene Stanley, and Walter Quattrociocchi. "The spreading of misinformation online." Proceedings of the National Academy of Sciences 113, no. 3 (2016): 554-559.

Ellsberg, D. (1961), "Risk, ambiguity, and the savage axioms", *Quarterly Journal of Economics* 75: 643-669

Gandal, Neil, J. T. Hamrick, Tyler Moore, and Tali Oberman. "Price manipulation in the Bitcoin ecosystem." *Journal of Monetary Economics* 95 (2018): 86-96.

Garcia, David, and Frank Schweitzer. (2015), "Social signals and algorithmic trading of Bitcoin." *Royal Society open science* 2, no. 9: 150288.

Gottfried, Jeffrey, and Elisa Shearer. 2016. "News Use across Social Media Platforms 2016." Pew Research Center, May 26. http://www. journalism.org/2016/05/26/news-use-acrosssocial-media-platforms-2016.

Hansen, LarsPeter, and Thomas J. Sargent. "Robust control and model uncertainty." *American Economic Review* 91, no. 2 (2001): 60-66.

Maccheroni, Fabio, Massimo Marinacci, and Aldo Rustichini. "Ambiguity aversion, robust ness, and the variational representation of preferences." *Econometrica* 74, no. 6 (2006): 1447-1498.

Mocanu D, Rossi L, Zhang Q, Karsai M, Quattrociocchi W (2015) Collective attention in the age of (mis) information. Comput Human Behav 51:1198–1204.

Nakamoto, Satoshi. "Bitcoin: A peer-to-peer electronic cash system." (2008).

Nguyen, Thien Hai, Kiyoaki Shirai, and Julien Velcin. "Sentiment analysis on social media for stock movement prediction." *Expert Systems with Applications* 42, no. 24 (2015): 9603-9611.

Von Neumann, John. "Morgenstern, 0." Theory of Games and Economic behavior 3 (1944).

Ranco, Gabriele, Darko Aleksovski, Guido Caldarelli, Miha Grčar, and Igor Mozetič. "The effects of Twitter sentiment on stock price returns." *PloS one* 10, no. 9 (2015): e0138441.

Savage, L. (1964), *The Foundations of Statistics* (Wiley, New York).

Silverman, Craig. 2016. "This Analysis Shows how Fake Election News Stories Outperformed Real News on Facebook." BuzzFeed News, November 16.