Distribution Agreement

Filtered News Sentiments and US Equities During the COVID-19 Pandemic

By

James Traynelis

Dr. Mi Luo

Adviser

Economics

Dr. Mi Luo

Adviser

Dr. Paloma Lopez de mesa Moyano

Committee Member

Dr. Jeff Rosensweig

Committee Member

2020

Filtered News Sentiments and US Equities During the COVID-19 Pandemic

By

James Traynelis

Dr. Mi Luo

Adviser

An abstract of
a thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Arts with Honors

Economics

2021

Abstract

Filtered News Sentiments and US Equities During the COVID-19 Pandemic

By James Traynelis

Narratives play a determinate role in financial markets. Financial news data can be used as a surrogate to quantify the spread and adoption of economic narratives. This paper looks at two narratives during the COVID-19 pandemic: one of extreme monetary and fiscal stimulus and another of a vaccine. Financial data from four US stock indices (NASDAQ, SPX, RUT, NBI) was regressed on VADER sentiment ratings from New York Times and Guardian headlines. Vaccine news sentiment tracked strongly with the NASDAQ Biotech Index and Russell 2000. Stimulus news sentiment, however, did not exhibit any explanatory power at the significance levels tested.

Filtered News Sentiments and US Equities During the COVID-19 Pandemic

By

James Traynelis

Dr. Mi Luo

Adviser

A thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Arts with Honors

Economics

2021

Acknowledgements

I am honored to thank the many people who made this thesis possible. It is difficult to overstate my gratitude to my thesis advisor and professor, Dr. Mi Luo, for her insights and enduring support throughout this project. I am outspokenly grateful to have had guidance from such a talented researcher as Dr. Luo. I would also like to thank the other committee members serving as the official examiners of my thesis, Dr. Jeff Rosensweig and Dr. Paloma Moyano, both for their time and motivation. This project would not be what it is today without the thoughtful feedback from each of these individuals. And lastly, I wish to briefly thank the many other professors who played a role by going above and beyond in their lectures to motivate the questions at the center of this research. To name a few, I am especially grateful of Professor Massoumi, Tom Smith, and David Fossati. The encouragement and patience of all these individuals has been empowering in ways I could not have foreseen. Thank you.

Contents

List of Figures

List of Tables

Filtered News Sentiments and US Equities During the COVID-19 Pandemic

James Traynelis

March 2021

1 Introduction

Economists have long sought to quantify components of human behavior that shape real economic outcomes. For centuries, behavioral economic theories were simply untestable without the rich qualitative, behavioral data economists have access today. Big data and modern computational methods have revolutionized the way that economists can quantify and study human behavior. Financial markets serve as a unique environment to study these behavioral phenomena. Markets are informationally efficient– and in general reflect an aggregate of institutional expectations based on all available information. However, there are nonetheless instances where prices may rise despite a contradicting state of reality. Such bubbles are a fascination for many, and have proved handsomely profitable for those who foresee them and absolutely devastating for those who do not. Understanding these distortions is essential to economists concerned with directing or advising policies. Using news and social media data, many behavioral economics have begun quantifying the psychological component behind many of these distortions.

The year 2020 is a unique case study for mass psychology and its ramifications in financial markets. The COVID-19 Pandemic induced worldwide lock-downs and economic contractions at a level not seen since the 2007 financial meltdown. US equity markets have endured both record volatility while also making new highs. The initial economic shock from the global pandemic and shutdown was followed by massive fiscal and economic stimulus measures and large-scale rapid vaccine development programs. Both responses became hot topics in a greater reflation narrative. Each facet in it's own right gives a compelling possible explanation for the rebound in US equities despite the persistent pandemic conditions. It wasn't long before the pending arrival of a vaccine and government stimulus became two popular narratives tracked by financiers and policy makers. Even US Federal Reserve Chairman Jerome Powell has cited these narratives claiming in a 2021 Federal Open Market Committee that the surge in asset prices has been largely driven by "expectations about vaccines" and not monetary policy [1]. The primary aim of this project is to investigate the interplay between these two narratives and financial markets over the course of the pandemic.

2 Literature Review

2.1 Narrative Economics

Humans have been using narratives to transmit ideas across time and space long before they even invented language. Narratives play a determinate role in the way humans understand reality and make forecasts about the future. One might be wondering, what do these narratives have to do with economics? Researchers have long studied the extent to which human psychology distorts financial markets. Narratives often serve as the basis for market distortions such as bubbles or over reactions.

One study in particular, carried out by Richard Thaler and his colleagues at The University of Chicago, examined investor overreactions in financial markets. Thaler measured investor behavior surrounding earnings announcements, specifically focusing on announcements where companies performed either considerably better or worse than expected, and then compared these movements to conventional pricing models like the Capital Asset Pricing Model (CAPM). Thaler concluded that, more often investors do, by in large, overreact to the release of new or unexpected information. Thaler and his colleagues also found that overreactions are asymmetrically larger when they result in losses [10]. This means that market inefficiencies not only exist but can have rather unconventional explanations such as human psychology.

Another behavioral economist Robert Shiller, in particular, is famous for his exploration into the "epidemiology" of economic narratives. By applying the Kermack-McKendrick SIR (Susceptible-Infected-Removed) model, Shiller observed that for certain economic narratives, just like diseases, their spread and transmission increased exponentially before eventually declining over time [30]. Shiller discovered that the 1920-21 recession was linked to a consumer boycott of goods. By counting the historical appearances of the word "Profiteer" in news articles and books, Shiller discovered a prevalent belief that corporations were profiteers and overcharging on goods. In his book, *Narrative Economics*, Shiller delves into how powerful narratives can be with other historical examples from The Great Depression and more recent 2008 financial meltdown [30]. Since its initial appearance and eventual diffusion across the globe, the COVID-19 pandemic has given rise to many narratives. The 2020 reflation in US equities despite persistent lock-downs and low economic output suggests there were additional narratives driving asset prices.

2.2 COVID-19 Pandemic

As briefly mentioned earlier, the year 2020 has been a wild year for financial markets. US equities were particularly hit hard by the COVID-19 pandemic. Stock indices such as the S&P 500 and NASDAQ posted their greatest losses since the 2008 Financial Crisis in late February 2020, dropping over 28% between February 17$^{\text{th}}$ and March 16$^{\text{th}}$. There was

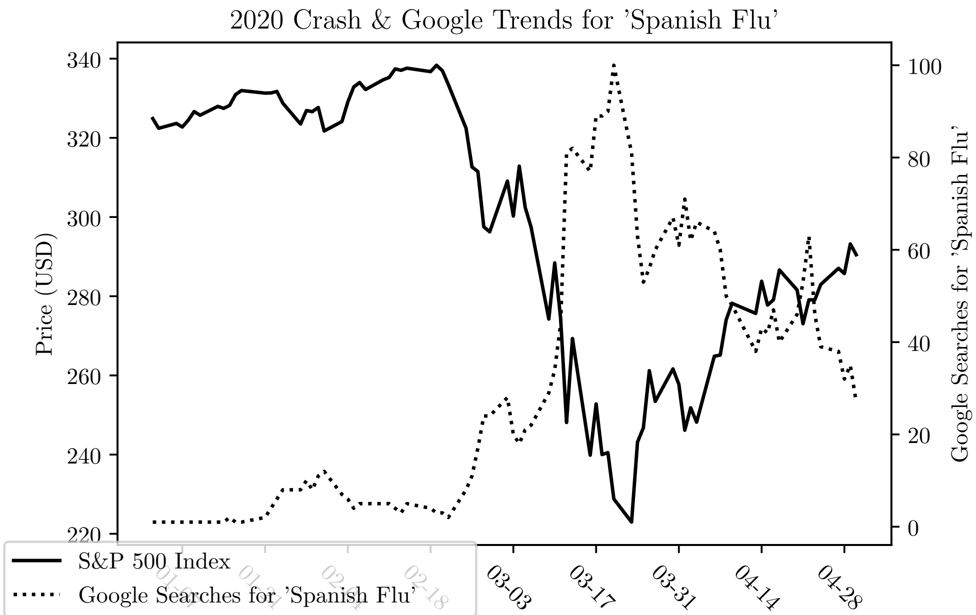

Figure 1: One Prevalent Narrative During the March 2020 Crash

a considerable time-lag between the virus' initial emergence and its consequent impact on equity prices. Diagnosing the direct catalyst of this sell off is difficult, but it was at least partially caused by the uncertainty surrounding the ensuing pandemic and virus itself. One interesting characteristic of the sell-off was that it was speculative, and financial markets responded to the pandemic well-before economic data indicated any stress [9]. Because financial markets function as an aggregate of future expectations, they are keenly responsive to news driven events. One partial explanation of the sell-off might have been the spike in mainstream news articles reporting on the virus. Figure 1 shows the 2020 crash graphed alongside google trends for the term "Spanish Flu". The widespread belief at the time that the COVID-19 Pandemic would play out similarly to the 1918 pandemic is exactly the kind of psychology described by Thaler and Shiller that may have been behind the crash.

While the vaccine narrative investigated in this paper is hyper-specific to the context on the 2020 COVID-19 Pandemic, there are several more-general reasons that monetary and fiscal stimulus might be linked to rebounds in equity prices following crises. Monetary stimulus describes a set of rather blunt tools that central banks can use to control the direction of market outcomes by easing or restricting the overall flow of credit in the economy. Several empirical papers have established there is a strong connection between monetary stimulus and asset prices[23, 7]. During 2020, the Federal Reserve's swift actions in response to the March 2020 liquidity crisis included the large-scale purchase of assets among several other measures. These purchases, or quantitative easing, effectively drives rates down. On one hand, this forces investors out of safe haven investments like money market accounts, treasuries, or cooperate bonds and into riskier investments with higher returns[23]. Alternatively, fiscal stimulus works by increasing consumption, which also aids the flow of money in an economy [3]. Both forms of stimulus have implications in financial markets and play a key role in economic recoveries.

2.3 Sensationalist News Reporting

Because the narratives examined in this paper originated during the midst of a public health crisis, they are uniquely different than the economic narratives studied by Shiller and his colleagues, which tend to be more grounded in economic fundamentals. News organs have a history of covering health emergencies with a sensationalist spin. The paramount 21^{st} century example of sensationalist reporting took place around the 2014 surges in Ebola epidemics across Africa. The 2014 Ebola outbreak made headlines worldwide, despite only one single death being outside of Africa[24]. Public health researchers have uncovered numerous inconsistencies in the way publishers have reported on other outbreaks and pandemics and it's likely that the most-recent pandemic is no exception [20].

The literature suggests that inaccuracies in health crises' news coverage have been increasing since the early 2000's [24]. For example, the 2009 H1N1 pandemic generated over twice the volume of coverage than the 2002 SARS pandemic [24, 20]. Considering the 2020 pandemic was a far more severe pandemic in terms of both infection and mortality rates, its safe to assume this trend identified by public health researchers has continued. Evaluating how the over-reporting or under-reporting of these health crises alters how people perceive them is vital to also understanding how they may play a role as narratives.

An interesting 2016 study published in *The Journal of Clinical Epidemiology* took and unique approach to quantifying the degree of sensationalism in news headlines. Researchers applied maximum entropy modeling to 163,433 articles mentioning the SARS and H1N1 pandemics, then manually rated the "Scientific quality" of each article using a 20-year-old scientific standard published in the early 90's[20, 27]. For an article to demonstatrate scientific quality, it needed score highly in seven key categories: Applicability, Opinions vs Facts, Validity, Magnitude, Precision, Consistency, and Consequences [27]. The results after comparing the hand-rated results and model were astonishing. Not only did the articles' median ratings fail to meet scientific objectivity standards, but the automated model replicated these results with 74% accuracy [20]. Sensationalism is overwhelmingly present in in the reporting of health crises, and can be very easily identified using automated modeling techniques.

Financial news is, by no exception, susceptible to making sensationalist appeals and over-reporting on hot topics, since after all, their revenue streams too rely on maximizing viewership. As price movements become more extreme, so too does the tendency for establishments to over-report [29]. Financial news also can be speculative or rumor-based, and not always credible. Since after all, a common mantra among investors is to "buy the rumor, sell the news". Rumors intimately shape public opinion and are infamous for causing widespread panic and instability [21]. In the media age, such rumors spread arguably faster than ever through news outlets, micro-blogs, to social media posts [32]. The dynamic role of this modern data in financial markets demands examination.

Sensationalism is by no means the only distortion in today's news outlets. A wide range of biases shape and change an article's reception of news events. The specific phrasing of headlines determines how readers will interpret and contextualize the article. Lewandowsky and his colleagues cite the example of taxes to illustrate how biases are implicitly contained in the specific phrasing of headlines. Consider how the “tax burden” or “civilian surcharge” carry immediate, implicit biases [25]. It is well proven that the people gravitate towards news stories that affirm their beliefs [4, 30]. This psychological pattern adds to the self-reinforcing nature of narratives. Narratives are multi-faceted, and may asymmetrically interact with markets– benefiting certain sectors more than others. Together, appeals to sensationalism and confirmation biases fuel irrational price movements that can sometimes result in market distortions.

2.4 News & Algorithmic Trading

The analysis of financial news with regards to stocks has become widely popular among researchers and practitioners over the last decade. With the robust collection of Natural Language Processing libraries that exist today, it’s no surprise that this modern approach has gained attention. For financial markets, the flow of information is vital. A constant stream of information feeds markets while simultaneously providing a trail of rich text data ripe for analysis. Natural Language Processing tools serve as very powerful ways to mine this stream of textual information for valuable insights and patterns.

For the most part, the a majority of the literature approaches the challenge of processing news articles as a large-scale regression problem, concerned with prioritizing fit. A large proportion of this literature arises from the field of computer science, where researchers are concerned with developing algorithmic trading strategies and the computational intricacies rather than understanding the economic factors driving markets. This project differs in the sense that its primary focus is applying these strategies to gain a deeper economic comprehension of the features influencing market outcomes.

As these Natural Language Processing methods are relatively new and evolving, working with text data comes with its own set of costs and limitations. Several challenges arise when working with text data. The natural flow of news tends to be extremely noisy. Over two million articles are published every day on hundreds of topics. Not only is the quantity of news immense, its speed of dissemination increases exponentially over time [14]. For these reasons, irrelevant or uncredible articles can lead to false signals. This makes it difficult to link a specific trend or narrative to a stock increase. For this same reason, most researchers have circumvented these problems by employing various filtering techniques [17, 16, 14]. It was shown by Challet in 2013 that random financial keywords used in a google trend's search turned out to be a good predictor of volatility in financial markets [8]. Filtering methods have increased precision across the literature. Although, in addition to understanding when exactly how the release of financial news is influencing outcomes in financial markets is extremely valuable to economists. And such is understanding the explanatory factors beneath such changes.

The impact of firm-specific articles on firms' respective stocks has been shown to be very strong in the literature [9, 14]. There have been many other attempts to further quantify this relationship such as in a study by Klussman and colleagues [14]. Filtering a set of articles to focus on a select asset or asset group, enhances the modelâs predictive accuracy greatly.

3 Methodology

3.1 Overview

This study aims to investigate whether narratives proliferated by large media outlets track strongly with price developments in financial markets. During the year 2020, a massive reflation in equity prices followed their sharp decline in response to the COVID-19 pandemic. There have been two narratives largely associated with this rebound in asset prices– one of extreme monetary and fiscal stimulus, another of an effective vaccine that would ultimately replace lock-downs as a combative measure against the disease. While both narratives are incredibly nuanced and display an extreme level of complexity, they still are propagated by an underlying, mass psychology that is increasingly quantifiable using various computational approaches. Because the transmission of ideas is remarkably similar to that of viral diseases, news headlines will be used as a proxy to quantitatively evaluate the degree of these narratives' dissemination [30]. To better understand how these ideas were transmitted on a global scale, and because the sheer volume of headlines published each day far exceeds the amount any human could read on their own, this paper will rely on Natural Language Processing Library VADER to evaluate each narrative's dissemination.

Similar approaches to sentiment analysis are wildly popular both in academic research and private enterprise. Natural Language Processing's popularity has only been magnified by the increasingly-accessible, digital stores of historical headlines and social media posts. Moreover, in 2018, a study estimated that nearly 70% of trading on US exchanges was algorithmic [18]. This number is only expected to have grown in the last three years. Automated, artificial processing of financial news is now a viable alternative approach to manually reading news articles. Social media has also transformed the way people consume news and, generally, information. According to another study conducted in 2018, 55% of surveyed adults over 30 reported getting their news from social media, which was twice as many as the number reported in 2010 [22]. In consideration of these trends, it is crucial to establish i) how well popular NLP libraries extract sentiment from news headlines ii) And secondly How the two narratives compete as explanations for the rebound in stock prices. This will be done by employing economic and computational models to analyze thousands of filtered headlines. All code and data used during this study will be freely accessible through Github at github.com/jvtraynelis/News-Sentiment-and-Stocks—2021-Senior-Thesis.

3.2 Data

3.2.1 News Data

Headlines from the New York Times and Guardian were selected as the main data source in this project for several reasons. First and foremost, both publishers provide public API's with both high call limits and very extensive online documentation. While more robust, subscription API's are available and often used by large commercial enterprises for similar purposes, both the New York Times and Guardian are reliable world class new sources with millions of readers worldwide. Their readers reflect a finance-savvy, college educated subset of the population who is probabilistic more likely to be involved in financial markets. Over 56% of The New York Times' readers and 65% of Guardian readers both hold a college degree or higher, which does not include those currently pursuing a degree [22]. 38% of The New York Times' readers earn at least $75,000 per year while Guardian readers were found to be 24% more likely to hold a mortgage and 32% more likely to have stocks and shares than the average UK citizen. It was also reported that 68% of Guardian readers are over the age of 35. Aside from these demographics being slightly skewed towards being more financially well-off and older, the gender breakdown is very balanced. 52% of New York Times' readers are male while 48% are female. Similarly, around 56% of Guardian readers are male while 44% are female. These statistics suggest that the subscribers to these news sources may be more inclined to follow international, financial news. Given these demographic characteristics, it's reasonable to conclude each source may cater to its educated base by publishing news stories in line with the common narratives propagated among these groups.

Financial news headlines are an ideal medium to study the transmission of economic narratives due to their concise delivery of complicated economic developments encoded in simple phrases and keywords. Headlines are also a unique target for sentiment analysis and classification because they tend to adhere to professional standards and are fairly uniform in size and term usage. Unlike tweets, headlines are carefully crafted and controlled to match their networks' presiding opinions. The resulting strings of words tend to be succinct, while still reflecting an aggregate of professional viewpoints from editors and journalists, whereas tweets, or other social media posts, tend to solely reflect their respective author's. Tweets are also far more susceptible to taking extreme or radical viewpoints that do not align with any prevailing paradigm. Perhaps an even more-compelling reason for the use of headlines, however, is their tendency to embody a degree of sensationalism. Because publishers rely on total impressions for their revenue, headlines are often designed to be eye-catching, perceived as relevant, and align with a prevailing ideology that suits their patronage. This results in frequent usage of buzzwords and other time-relevant phrases that quickly catch the eye of prospective readers. As a direct result, headlines exhibit extremely strong semantic patterns when grouped by topics and charted over time– thereby facilitating the perfect environment for an in depth analysis of economic narratives.

To ensure a sufficient sample size was acquired, for both vaccine and stimulus-related headlines, API requests were made collecting historical data over a 15 month period beginning on January 1st 2020 and ending on March 5th, 2021. March 5th was chosen as the cut-off date because it represents a significant inflection point in both vaccine development and monetary stimulus measures. On March 5th Moderna announced its pursuit of a COVID-19 vaccine. Shortly thereafter, on March 15th, 2020 in response to the pandemic-devastated financial markets, the US Federal Reserve announced emergency large-scale easing programs aimed at preserving the fragile flow of credit in the US economy. This included cutting the Federal Funds Rate to zero, quantitative easing, relaxing certain capital requirements, and several other measures to resuscitate a contracting economy. Early March marked a potential starting date for the reflation narratives, and will be further referred to as such.

API's, or application programming interfaces, are fantastic ways to collect both historical and real-time data. The high institutional demand, however, creates a very lucrative market for them, which can present challenges to unfunded developers. Initially, despite most of the later analysis being completed using Python 3.9.2, R 4.0.5 was used to facilitate all

API calls using the jsonlite package. The data retrieved via API calls was filtered via built-in query endpoints to subset headlines only relating to vaccines or stimulus news. Each query endpoint returned articles so long as they contain the specified query keyword in the article's body or headline. Vaccine and stimulus headlines were gathered separately to facilitate a later comparison between the two sets of headlines. Although both API's used in this project are free, to prevent the abuse of their interfaces, the New York Times sets a limit of 4,000 requests per day, whereas the Guardian limits requests 5,000 per day. In order to avoid exceeding these call limits, data was methodically collected over 8 days. Each API conveniently returns data in chronological order, thus permitting the continuous retrieval of headlines over several days. 13,780 headlines were gathered in total between January 1$^{\text{st}}$, 2020 and March 5$^{\text{th}}$, 2021. 5,931 of these headlines were stimulus-related, while 7,849 belonged to the vaccine-relevant data-set. Each resulting data set contained article headlines and the date and time they were published. Among the two groups of headlines, there were strong semantic patterns that reflected the target themes. Table 1 displays five randomly selected headlines from each data-set.

Vaccine-related Headlines
US pharma company raises vaccine hopes but more trials are vital, say experts
AstraZeneca, Under Fire for Vaccine Safety, Releases Trial Blueprints
China Approves Covid-19 Vaccine as It Moves to Inoculate Millions
New Pfizer Results: Coronavirus Vaccine Is Safe and 95% Effective
Russia says data on Sputnik Covid vaccine shows 95% efficacy
Stimulus-related Headlines
Unemployment Claims Remain High as Millions Still Struggle to Find Work
Stimulus Deal Falters as McConnell Signals Republican Resistance
Congress Passes Huge Coronavirus Relief Bill
Put the Money Printer on Autopilot
At Long Last, a Stimulus Nears

Table 1: Sample Headlines Gathered from The New York Times & Guardian API's

3.2.2 Financial Data

While the larger reflation narrative has been widely attributed to extreme monetary stimulus and the rapid development of a COVID-19 vaccine. Certain sectors in the economy have rebounded asymmetrically well compared to others such as small businesses or other contact-based businesses worst affected by the nation-wide lock-downs. To better understand these asymmetries, it is important to understand the degree to which these narratives played in determining market outcomes in each sector. Thus four stock indices were selected and examined, each chosen to represent different sectors of the economy. The following were chosen:

1. NASDAQ: Has a historical record of tracking with the US economy since it was founded in 1971.
2. S&P 500: Contains 500 stocks, chosen based on market size, liquidity, and industry, which together represent leading industries within the U.S. economy.
3. Russell 2000 : Reflects 2,000 of the smallest publicly-traded U.S. companies based on market capitalization.
4. NASDAQ Biotech Index: Contains NASDAQ-listed biotech and pharmaceutical companies.

Historical stock data was downloaded from Yahoo Finance for the same date ranges as headlines (March 5^{th}, 2019 - March 5^{th} 2021). The data included date, open price, high, low, closing price, adjusted closing price and trading volume. Only date, opening price, adjusted price, and volume were saved for analysis, while high, low and regular closing price were ignored. The adjusted closing price was used instead of the regular close as a more accurate pricing strategy. The adjusted close is essentially the closing price after all cash/stock dividends and stock splits have been taken into account. Because all assets used were indices comprised of hundreds of individual stocks, it was crucial to use the adjusted close as the true price of the index after these dividends and splits had been factored out.

3.3 Data Filtering

3.3.1 Dates

The scope of this study was limited to observing longer-term price developments in context of the larger, macro-economic picture, and thus drastically less concerned about capturing the high-frequency, intraday price movements following news releases. The dynamics of news releases and high-frequency price responses have, however, been covered extensively in the literature, particularly by economists Busse and Green in the early 2000's [6, 13]. Both studies found that asset prices respond in a matter of seconds to critical news releases. For this project, however, daily frequencies were sufficient to observe the amassed reaction to news releases across several indices and macroeconomic implications. However, initially, because markets are not open on weekends or during national holidays, news and stock data did not have the same frequencies. This was because all four indices are based in the United States and do not trade on weekends or national holidays. News data on the other hand, flows unrestricted, everyday, 24 hours a day. In the data obtained, thousands of articles were either published outside of trading hours or on days on which financial markets were closed (i.e National holidays or weekends). It would be foolish to discount any news simply for being published outside of trading hours. Thus a new date was computed for each article that corresponded to the next trading day– following Busse's conclusion that price responds almost instantaneously to news [6]. This new date was only used if an article was published outside of regular trading hours. So if a headline was published stating Astrazeneca vaccine trials were halted on a Tuesday evening at 8pm for instance, that headline would be assigned a new date corresponding to the following day, with the assumption that it would impact markets after open the next morning. For headlines published on Saturday or Sunday, their date was adjusted to match the following Monday's date. The same method was applied to the holidays: New Year's Day, MLK Jr. Day, Presidents Day, Memorial Day, Independence Day, Labor Day, Thanksgiving Day, and Christmas. Several other steps were taken before the data was merged by the new date to the corresponding financial data. While R was used extensively in the scraping and cleaning of the data, python was used during the remainder of the project because of it's powerful Natural Language Processing libraries.

3.3.2 Sentiment Analysis

Sentiment analysis has exploded over the past ten years in popularity, and is now frequently cited in the literature. What is meant by sentiment? Before proceeding, it is important to explain what sentiment is and how it could relate to prices. Investor sentiment can broadly range from and expectation of belief about the future price and risks of an asset that in not justified by the current set of facts [5]. When it comes to text-sentiment libraries, there are countless, highly-sophisticated libraries to choose from. Because headlines tend to be short in length, the sentiment analysis tool VADER (Valence Aware Dictionary and sEntiment Reasoner) was used to score the rating of each headline. VADER is one of the more powerful Natural Language Processing libraries, and is trained on millions of short isolated texts. The library is also very commonly implemented in algorithmic trading strategies for its speed and versatility[15]. VADER is designed to measure both the intensity and sentiment of short, concise sentences. The mathematical derivations of sentiment scores will, however, be further discussed in the time-series analysis section of this methodology. It is most commonly used in analyzing tweets and online reviews, both of which tend to be comparable in length to news headlines. This characteristic is often cited as a reason why VADER also performs well on news headlines. And similar applications have been well-documented in the literature[15, 29]. While a large fraction of the literature on the subject favors automated, machine learning models, there is still considerable insight that can be gained from using simpler, descriptive approaches to sentiment analysis. These techniques are elegantly laid out by Guidi in his

2017 paper on tweets and stock price prediction [15].

VADER measures the intensity and polarity of texts. It provides three polarity scores: A positive, negative, and neutral score where neutral is the complement to the sum of negative and positive scores. Each score is constrained between -4 and 4. VADER then computes a compound score which is the normalized sum of its positive and negative scores, shown in equation 2, where x is the sum of valence scores.

$$x = VADER_{positive} + VADER_{negative} \tag{1}$$

The default normalization constant $\alpha = 15$ was used in this experiment, as recommended by the package. Thus the normalization equation applied in this experiment is shown in equation 2.

$$x = \frac{x}{\sqrt{x^2 + 15}} \tag{2}$$

Table 2 shows VADER's compound scoring system in action on a sample of 8 headlines. While VADER extracts some of the relevant sentiment, it is immediately apparent VADER has a limited comprehension of the complicated pandemic context. However, VADER does do an exceptional job identifying the intensity and polarity of vaccine headlines.

Although using the compound normalized sum of polarities is a common and easy way to obtain a snapshot of the overall sentiment in a given time period, it would be insufficient to discard the individual negative and positive ratings. This is for the simple reason that what might be good news for one index is not necessarily good for another. The set of stock indices chosen requires we consider the partial effects of both negative and positive news in addition to the normalized sum of the two. That said, neutral scores are also relevant to this experiment, perhaps best understood as a measure of no news. By considering each of VADER's negative, positive, compound, and neutral scores different hypotheses can be tested. For instance, is no news good news? Is negative vaccine news worse for biotech stocks? Does positive sentiment around monetary and fiscal stimulus lead to price increases?

Neutral:
Coronavirus Variant Is Indeed More Transmissible, New Study Suggests Compound Rating: 0.0 Becoming a Part of the Story Compound Rating: 0.0
Positive:
New Pfizer Results: Coronavirus Vaccine Is Safe and 95% Effective Compound Rating: 0.718 The surge in coronavirus cases is helping speed up the development of vaccines. Compound Rating: 0.296 A day of hope, and warning, as a vaccine is rolled out in the U.S. Compound Rating: 0.128
Negative:
How Bad Will the Coronavirus Outbreak Get? Here Are 6 Key Factors Compound Rating: -0.542 America and the Virus: 'A Colossal Failure of Leadership' Compound Rating: -0.511 Russia Is Slow to Administer Virus Vaccine Despite Kremlin's Approval Compound Rating: -0.372

Table 2: VADER Compound Sentiment Scores

3.4 Textual Analysis

Removing stop words and lemmatization are two popular methods for extracting the significant words in text. To get an understanding of the text data and the common trends between narratives, the text data was first filtered. Stop words were removed and stemming was conducted to consolidate important terms that occurred most frequently in the data. Stop-words are the often-meaningless, yet frequent words that show up in text data (i.e. the, he, is, and etc.). Lemmatization reduces the amount of unique words by extracting the root term. For example, the two terms "vaccines" and "vaccinate" both refer to the same root "vaccin" but without first stemming will be counted as two separate terms. The Lancaster stemming method, included in the Natural Language Took Kit Library (NLTK), is one of the more aggressive and was used in this textual analysis to group terms based on root.

Table 3 shows the top-15 most frequent words in each of the two data sets after removing stop-words and stemming. The emphasized terms are uniquely more common to that narrative. Among the stimulus headlines, the roots "econom", "stimul", "busy", and "market" all appear in the top 15 most common roots. This proves that the filtering during data collection was successful in collecting headlines relevant to the targeted narrative. Similar patterns can be identified in the vaccine news data, where roots "vaccin", "cas" (as in cases), "lockdown", and "test" (as in testing) all appear in the top-15 most frequent terms. It's important to point out a few anomalies unique to this data set. You'll notice that the root "hap" appears in the top-15 terms for both news groupings. The root "hap" actually corresponds to the term happened. This is because both the New York Times and Guardian publishes several recurring, weekly summaries called "As it Happened?" and "How it happened". It's crucial to point out that these two lists do not provide any context relevant to the financial data, yet table 3 still provides a useful overview of the semantic patterns in the news data.

Stimulus News		Vaccine News	
Word	Frequency	Word	Frequency
coronavir	821	coronavir	1261
hap	576	covid	856
econom	**452**	hap	843
trump	446	**vaccin**	**750**
covid	291	trump	438
brief	265	uk	356
stimul	**259**	us	338
fed	**255**	new	326
us	246	**cas**	**311**
new	244	brief	305
busy	**232**	say	298
say	224	**lockdown**	**220**
market	**175**	**test**	**167**
vir	**169**	pandem	158
pandem	168	**austral**	**157**

Table 3: Top 15 Most Frequent Words

3.5 Time Series Analysis

3.5.1 Detrending Time Series Stock Data

Because the focus of this project is primarily on determining which narrative has more explanatory power and not on predicting specific prices, the historical price data was detrended to maximize the relevant signal. Time series data poses several challenges, one of which is noise or cyclical trends that tend to obscure the underlying signal from being studied. The non-stationary, financial time-series data in this project were very noisy. 2020 was after all a year with record volatility levels. Fortunately, there are well established methods for detrending economic variables with heavy cyclical components. One especially popular method is the Hodrick-Prescott filter. This detrending method effectively decomposes a time series y_t into a cyclical component ζ_t and trend τ_t, shown below [19].

$$y_t = \zeta_t + \tau_t, \quad t = 1, 2, ..., T \tag{3}$$

The components are derived by minimizing the quadratic loss function in equation 4, which minimizes the predictive accuracy lost from the model. The first term in the equation minimizes the variance of ζ_t, and coefficient λ, or smoothing factor. The function simultaneously punishes a lack of smoothness in the second term. It's important to note that as λ converges to zero, the trend τ_t becomes equal to the original time series y_t. Conversely, as λ diverges to ∞, τ_t becomes linear.

$$\min_{\tau_t} \sum_{t}^{T} \zeta_t^2 + \lambda \sum_{t=1}^{T} [(\tau_t - \tau_{t-1}) - (\tau_{t-1} - \tau_{t-2})]^2 \tag{4}$$

Traditionally a smoothing factor $\lambda = 1600$ is reserved for macro variables with quarterly data [26]. Because the stock market contains cyclical patterns similar to those of other macro variables, the Hodrick-Prescott filter is still a viable approach using the smoothing factor $\lambda = 1600$. Stock data tends to have a larger cyclical component than the business cycle too, which further justifies the use of such a low smoothing factor [2]. For these reasons, not only can the Hodrick-Prescott filter be applied, but also does an excellent job of removing the noise and cyclical component while preserving the underlying signal. Figure 2 shows this detrending in action for the S&P 500. The same procedure and smoothing factor was used on all indices, the S&P 500 is just one visual example. In figure 2, the signal extracted retains certain features of the original series, yet is also dramatically smoother. Using this filtered time series, the financial news immediately had stronger relationships with the news sentiment scores.

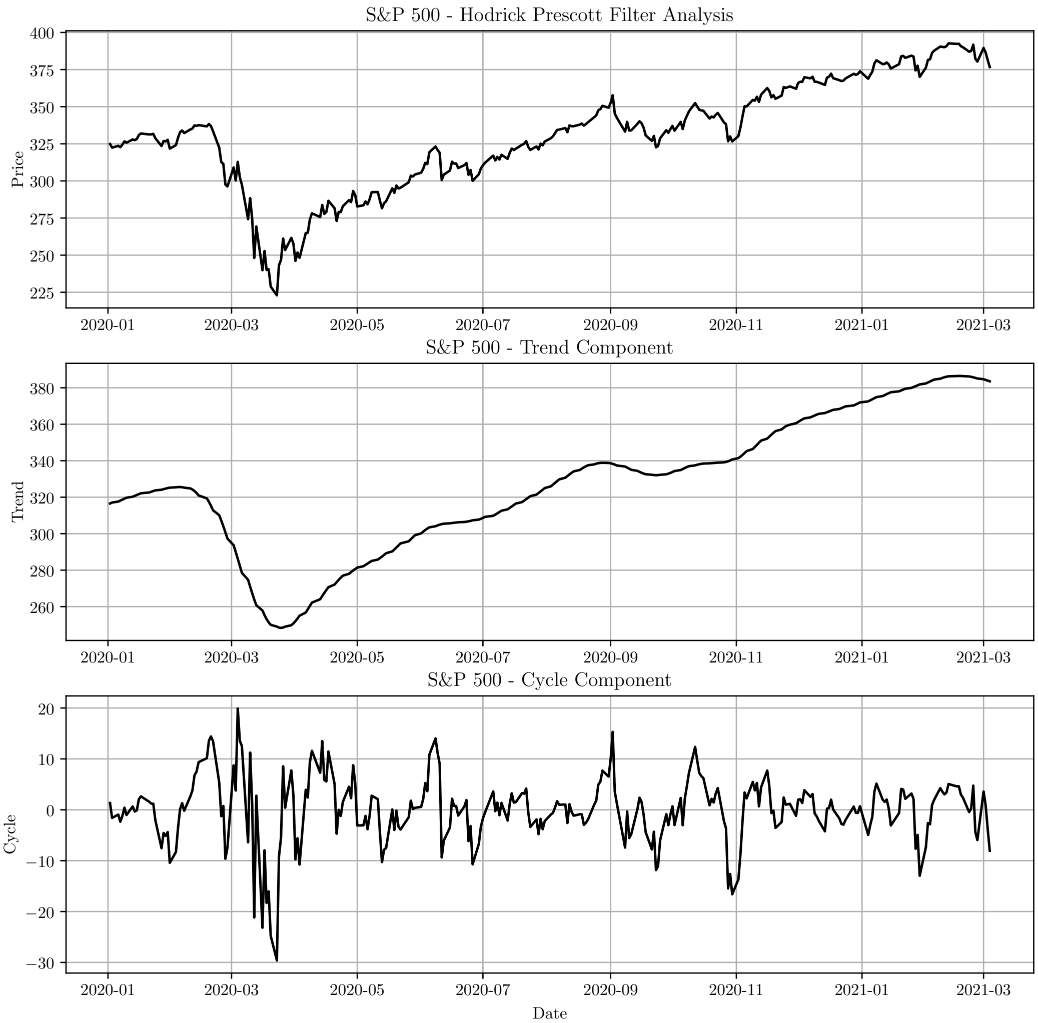

Figure 2: Detrending Financial Data

3.5.2 The Dependent Variable

Because each financial time series data was non-stationary and had its own unique mean and variance, the dependent variable needed to be something comparable between assets.

In alignment with Atkins' work predicting stock volatility and price using financial news, in which he concluded news data to be a very poor predictor of closing price, this study will take a similar approach and focus on predicting directional movement. A discrete binary variable indicating if the stock increased on a given day, was calculated. Denoted by $P(Asset\ Increased)$, the dependent variable was equal to 1 if the stock did increase and 0 if not. This variable was based on the percent change calculated using the open and adjusted closing prices during regular trading hours after detrending the time series. If the stock price closed higher than it opened, after dividends and splits had been filtered out, then $P(Asset\ Increased = 1)$. Logistic regressions have interesting interpretations, namely the dependent variable becomes a probability distribution function that follows a Bernoulli distribution with probability p_t. One implication of using a linear probability model, however, is that the Gauss-Markov assumption of homoskedasticity quickly fails. Yet logistic models are still quite powerful way to study binary outcomes. The conditional expectation and conditional probability can then be written as

$$E_{t-1}(y_t) = P_{t-1}(y_t = 1) = p_t \tag{5}$$

After computing this discrete variable, indices were compared based on proportion of increases to decreases. Figure 3 shows the distribution of days based on whether or not the index closed higher than it opened. This plot reveals key characteristics of each individual index that will be discussed later. Notice, however, that both the Russel 2000 and NASDAQ Biotech Index have considerably less up-days than their coutnerparties.

3.5.3 Cointegration

Spurious correlation is a common issue encountered in econometric models that deal with one time-series regressed on another. Spurious correlation often leads to incorrect causal interpretations from test statistics because either the two series are coincidentally correlated

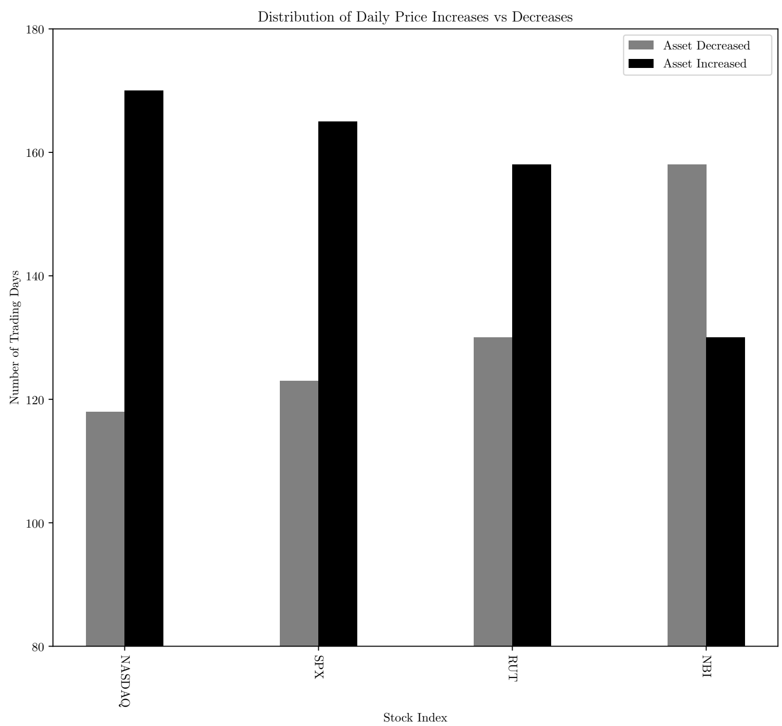

Figure 3: Distribution of Negative & Positive Trading Days Per Asset

over time or correlated with some unknown third variable. When this is the case, the two functions also tend to be stationary, only different by some factor β [31]. Thus, before proceeding to complete any regression, the time-series were tested for cointegration.

One popular test for cointegration is the the Augmented Engle-Granger Two-step Test. This testing method first introduced in 1987 works by applying the Augmented Dickey-Fuller (ADF) unit root test [12, 11]. The Engle-Granger method determines whether two times series variables are cointegrated by first obtaining the residuals from a simple linear regression, $\hat{u}_t$, then by carrying out an ADF root test. This test can very powerfully be augmented by including additional lags of $\hat{u}_t$ to account for any serial correlation, but will not be for this experiment. The Engle-Granger hypotheses are typically written as:

$$H_0 : \textit{No Cointegration Exists between } x \textit{ and } y$$

$$H_1 : \textit{Cointegration Exists between } x \textit{ and } y \ .$$

Conveniently, the Python library Stats Models includes a method for conducting Engle-Granger tests. The Stats Models method returns t-statistics and critical values from the ADF root test. Using this library, sentiment scores and the discrete variable $P(AssetIncreased)$ were checked for cointegration. Table 4 shows the results from the Engle-Granger test.

Index	Positive	Negative	Compound	Critical Values
NASDAQ	-2.82	-2.86	-2.85	-3.93, -3.36, -3.06
SPY	-3.97***	-3.47**	-3.41**	-3.93, -3.36, -3.06
RUT	-3.03	-2.94	-2.75	-3.93, -3.36, -3.06
NBI	-3.47**	-3.86**	-3.56**	-3.93, -3.36, -3.06
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 4: Engle-Granger Test for Cointegration

The dependent variable $P(Asset\,Increased)$ was cointegrated with sentiment for the S&P 500 and NASDAQ Biotech Index according to the results from the Engle-Granger test of cointegration. This cointegration is also visible in figure 4. There was also a considerable level of cointegration in the other indices, although not statistically significant at the 10% level. Thus, to err on the side of caution, the first difference in all four time series was taken to avoid possibly obtaining inflated test statistics due to any cointegration.

3.5.4 Taking First Difference

Because the two time series data are cointegrated, meaning they are upward or downward trending together, there is a possibility that the apparent correlation may be spurious or false. Differencing is a method of transforming a time-series data to rid of any cointegration with another time series. The first difference, or lag-1 difference, is when the difference is taken between each consecutive observation [31]. When the first difference of a binary variable is taken however, the dependent series is no longer binary but rather ranges from -1 to 1. Although, the dependent series' variances still fail the Gauss-Markov assumption of homoskedasticity [31]. Thus, in absence of this Gauss-Markov assumption, a feasible Generalized Least Squares (GLS) regression was used in placement of ordinary least Squares (OLS). Because the data has a time series dimension, the model used in the feasible GLS is defined as:

$$P_{t-1}(y_t = 1) = \alpha_0 + \beta_1 Pos_t + \beta_2 Neg_t + \beta_3 Neu_t + \beta_4 Comp_t + u_t \tag{6}$$

Where Pos_t, Neg_t, and Neu_t are the positive, negative, and neutral VADER sentiment ratings respectively and t is one business day. $P_{t-1}(y_t = 1)$ is a binary variable indicating whether the filtered trend increased during the period t. After taking the lag-1 difference, the model becomes

$$\Delta y_t = \theta P_{t-1}(y_t = 1) + e_t. \tag{7}$$

A feasible GLS regression was then used to estimate coefficients θ because it is the best linear unbiased estimate under the presence of heteroskedasticity [31].

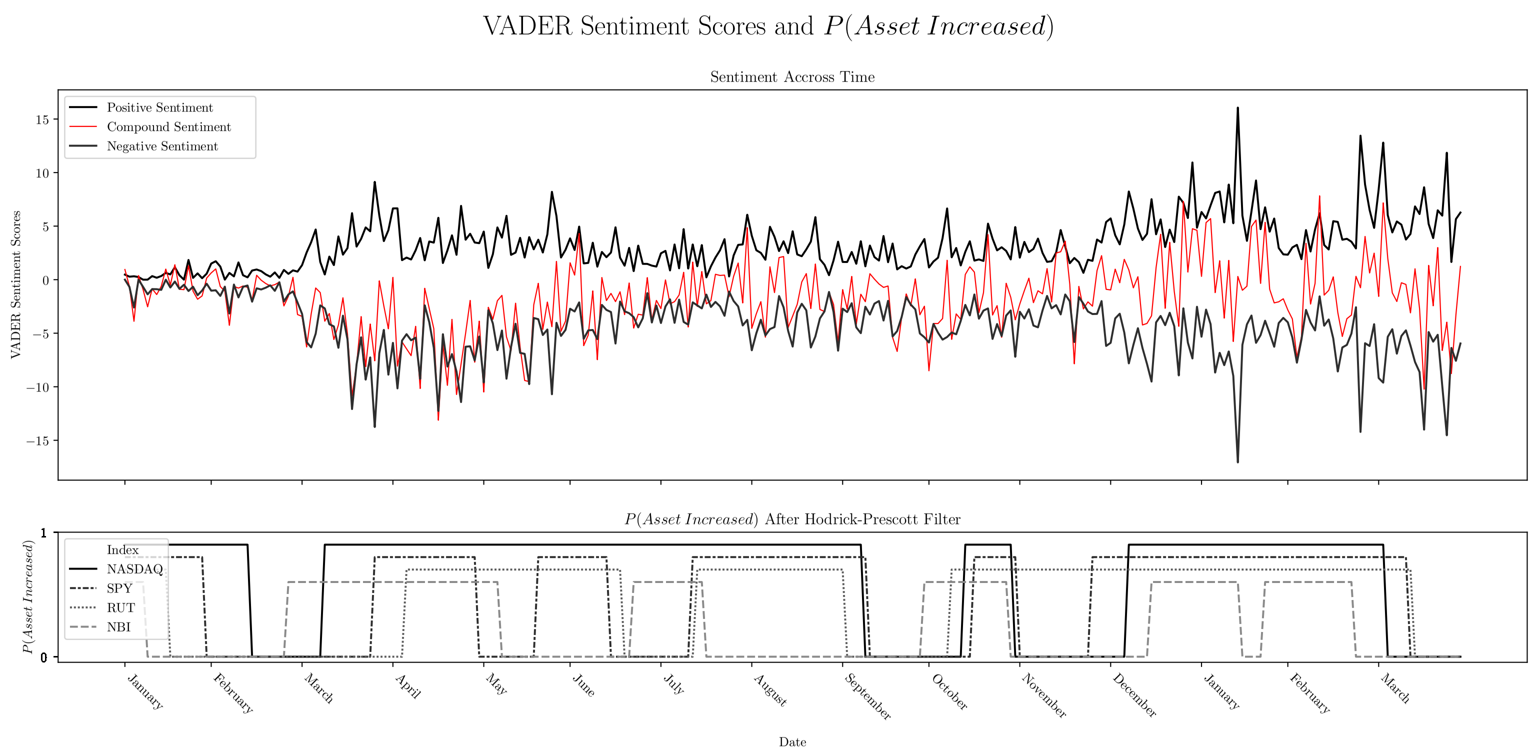

Figure 4: Sentiment Daily-Total Scores and the $P(AssetIncreased)$

4 Results

4.0.1 Overview

The results in this experiment align with other findings in the literature that sentiment derived from financial news does indeed track with financial markets. The larger aim of this experiment was rather to gain a better insight into how different narratives track with financial markets. More specifically, this paper looks at exactly how expectations about vaccines and stimulus throughout the COVID-19 Pandemic formed a larger reflation narrative that drove US equities to new highs despite the 2020 crash in response to the pandemic. Any causation would be difficult to establish with the limited news data, yet news headlines from this period can still serve as a useful surrogate to measure the contagion of these narratives. And while the data is certainly not perfect, the sentiment of these headlines was still a good indicator of directional returns. There was strong evidence suggesting vaccine sentiment extracted from the news data was correlated with the equity reflation between January 1$^{\text{st}}$, 2020 and March 5$^{\text{th}}$, 2021. However, stimulus-related sentiment did not exhibit any statistical significant explanatory power over the indices studied during this experiment. This might be due to imprecisions in VADER sentiment scores or the small sample of indices tested.

4.0.2 Vaccine Sentiment

	Dependent variable: P(Asset Increased)			
	NASDAQ	S&P 500	Russell 2000	NBI
Compound	-0.007	-0.001	-0.018	0.035**
	(0.013)	(0.016)	(0.012)	(0.015)
Intercept	-0.004	-0.003	-0.004	-0.004
	(0.009)	(0.012)	(0.009)	(0.011)
Negative	0.016	-0.003	0.041*	-0.066**
	(0.025)	(0.031)	(0.025)	(0.031)
Neutral	0.000	-0.002	0.003***	0.001
	(0.001)	(0.001)	(0.001)	(0.001)
Positive	0.021	0.009	0.009	-0.052*
	(0.024)	(0.030)	(0.024)	(0.030)
Observations	288	288	288	288
R^2	0.010	0.004	0.039	0.047
Adjusted R^2	-0.004	-0.010	0.026	0.033
Residual Std. Error	0.156	0.197	0.154	0.192
F Statistic	0.743	0.319	2.903**	3.452***
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 5: Feasible GLS on Vaccine Sentiments

Considering that the NASDAQ Biotech index contains the the Biotech & pharmaceutical companies listed on the NASDAQ, its correlation with vaccine sentiments makes resonable sense. As shown in table 5, only the Russel 2000 and NASDAQ Biotech Index show correlation to any of the Vaccine Sentiments.

These findings perhaps suggest that there was a third competing narrative or set of narratives that more closely tracked with financial markets during 2020. One potential candidate would be the outstanding performance of tech stocks during 2020. Big tech stocks like Apple, Microsoft, Amazon, Google, and Facebook make up 20% of the S&P 500. Table 6 illustrates the breakdown of weights between tech stocks in the S&P 500 [28]. As lock-downs spread too, some tech stocks such as online-conferencing platform Zoom grew over 600%. Because the S&P 500 and NASDAQ indices contain a sizeable fraction of large tech stocks that preformed exceptionally well during 2020, it's conceivable their prices did not track as closely with vaccine developments. This theory could be tested with a similar methodology using indices that contain primarily tech stocks. Perhaps "big-tech" could be used as a keyword for this additional narrative. More research should be conducted to test this hypothesis.

Company	Weight
Apple	6.79%
Microsoft	5.45%
Amazon	4.40%
Alphabet	3.28%
Facebook	2.07%
Total Share	19.92%

Table 6: Distribution of Tech Stocks in the S&P 500

4.0.3 Stimulus Sentiment

Narratives about massive monetary and fiscal stimulus absolutely characterizes the year 2020 for financial markets. On one hand, the Federal reserve expanded its balance sheet 4 trillion dollars. On the other, during both the Trump and Biden administrations, the government enacted expansionary fiscal policies not seen since the great recession and great depression eras. Despite these programs, there was no significant evidence suggesting stimulus related sentiment had explanatory power over any of the four indices. Additionally, after grouping all headlines together, both vaccine and stimulus headlines together, there was still no significant evidence suggesting stimulus sentiment had explanatory power over whether or not the indices increased. These results from the feasible GLS regressions could have one of two explanations: i) VADER, the Natural Language Processing library used, did not accurately classify the sentiment of stimulus headlines. ii) Or alternatively, Federal Reserve Chair, Jerome Powell, was correct in stating that vaccine expectations formed the prevailing narrative during 2020, and stimulus headlines did not track with equity prices.

See table 7 for more in-depth results.

	Dependent variable: $P(Asset\ Increased)$			
	NASDAQ	S&P 500	Russell 2000	NBI
Compound	0.004	0.004	-0.012	-0.007
	(0.012)	(0.015)	(0.012)	(0.015)
Intercept	-0.003	-0.004	-0.003	-0.004
	(0.009)	(0.011)	(0.009)	(0.011)
Negative	0.000	-0.009	0.019	0.017
	(0.021)	(0.026)	(0.021)	(0.026)
Neutral	0.001	0.003	-0.000	0.002
	(0.002)	(0.002)	(0.002)	(0.002)
Positive	-0.003	-0.005	0.016	0.017
	(0.021)	(0.027)	(0.021)	(0.027)
Observations	293	293	293	293
R^2	0.005	0.019	0.005	0.007
Adjusted R^2	-0.009	0.006	-0.009	-0.006
Residual Std. Error	0.156	0.194	0.155	0.195
F Statistic	0.326	1.420	0.343	0.544
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 7: Feasible GLS on Stimulus Sentiments

The weak correlations between both vaccine and stimulus headline sentiments and stock indices suggests there are great limitations to the data used in for this experiment. Thus, its very likely, some relationship may still exist and could be uncovered using a larger data set or revising the methodology to apply a more sophisticated sentiment analysis library. One reason for this being, headlines about the Federal Reserve's actions are summarized facetious or slightly more negatively. In other words, a reason for any inaccuracy may be VADER's inability to really understand which headlines are beneficial to the economy.

Alternatively, these results may genuinely indicate that stimulus news sentiment did not have explanatory power over whether or not the indices increased. It could be the case that stimulus news held predictive power over financial stock indices that might have been more directly impacted by monetary and fiscal stimulus. It's key to note here as well that fiscal and monetary stimulus may be less directly related to equities in general. There tends to be a much more direct relationship between economic stimulus and bond prices. on one hand, monetary policy's key weapon, rate targeting, tend to have well-pronounced effects in the bond market. Long-term interest rates play a major role in the term structure for bonds. When the US Federal Reserve announced in early March 2020 that it intended to keep rates close to zero and increased its purchase of long-term bonds, it effectively reduced term premiums. And while these changes to the term structure are the direct implication of the monetary stimulus measures, they may have a delayed effect on other asset classes such as those studied in this paper.

5 Conclusion

Narratives are convenient ways for people to communicate information over time and space. They should, however, be challenged, particularly when based on assumptions rather than hard-backed evidence. This study proved that narratives may not always be as explanatory as people would like to think. Stimulus news sentiment in this experiment, did not show a strong relationship with price movements on any of the four indices. Various filtering techniques explored in this project proved successful in obtaining a hyper-specific sub grouping of headlines. Vaccine News sentiment did however exhibit explanatory power for both the Russel 2000 and NASDAQ Biotech Index. The vaccine development narrative played a significant role in predicting directional change in price for the two indices worst affected by the pandemic. However more research should be conducted before making any causal inferences.

This study faced several limitations along the way, and like all research could serve as the foundation for more specific inquiries. Some future directions it may want to pursue are an extension of the data sets to include a larger sample of indices and larger set of headlines, and the incorporation of alternative sentiment analysis libraries. The sentiment analysis library, VADER, applied in this project did not offer as strong results as one might expect. and performed rather inconsistently when provided headlines from complicated financial environments. In the future, more research should be conducted to better understand how narratives drive asset prices.

References

[1] Chair powell's press conference. *Federal Reserve Media Center*, Jan 2021.

[2] Klaus Adam and Sebastian Merkel. Stock price cycles and business cycles. 2019.

[3] Luca Agnello and Ricardo M Sousa. Fiscal policy and asset prices. *Bulletin of Economic Research*, 65(2):154–177, 2013.

[4] Hunt Allcott and Matthew Gentzkow. Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, 31(2):211–36, 2017.

[5] Malcolm Baker and Jeffrey Wurgler. Investor sentiment in the stock market. *Journal of Economic Perspectives*, 21(2):129–152, 2007.

[6] Jeffrey A Busse and T Clifton Green. Market efficiency in real time. *Journal of Financial Economics*, 65(3):415–437, 2002.

[7] Indraneel Chakraborty, Itay Goldstein, and Andrew MacKinlay. Monetary stimulus and bank lending. *Journal of Financial Economics*, 136(1):189–218, 2020.

[8] Damien Challet and Ahmed Bel Hadj Ayed. Predicting financial markets with google trends and not so random keywords. *SSRN*, 2013.

[9] Anna Cieslak and Hao Pang. Common shocks in stocks and bonds. 2020.

[10] Werner FM De Bondt and Richard Thaler. Does the stock market overreact? *The Journal of Finance*, 40(3):793–805, 1985.

[11] James Durbin and Geoffrey S Watson. Testing for serial correlation in least squares regression: I. *Biometrika*, 37(3/4):409–428, 1950.

[12] Robert F Engle and Clive WJ Granger. Co-integration and error correction: representation, estimation, and testing. *Econometrica: journal of the Econometric Society*, pages 251–276, 1987.

[13] T Clifton Green. Economic news and the impact of trading on bond prices. *The Journal of Finance*, 59(3):1201–1233, 2004.

[14] Axel Groß-Klußmann and Nikolaus Hautsch. When machines read the news: Using automated text analytics to quantify high frequency news-implied market reactions. *Journal of Empirical Finance*, 18(2):321–340, 2011.

[15] Mirco Guidi. Extracting information from social media to track financial markets. B.S. thesis, Università Ca'Foscari Venezia, 2017.

[16] Michael Hagenau, Michael Liebmann, and Dirk Neumann. Automated news reading: Stock price prediction based on financial news using context-capturing features. *Decision Support Systems*, 55(3):685–697, 2013.

[17] Steven L Heston and Nitish Ranjan Sinha. News vs. sentiment: Predicting stock returns from news stories. *Financial Analysts Journal*, 73(3):67–83, 2017.

[18] Martin Hilbert and David Darmon. How complexity and uncertainty grew with algorithmic trading. *Entropy*, 22(5):499, Apr 2020.

[19] Robert J Hodrick and Edward C Prescott. Postwar us business cycles: an empirical investigation. *Journal of Money, credit, and Banking*, pages 1–16, 1997.

[20] Steven J Hoffman and Victoria Justicz. Automatically quantifying the scientific quality and sensationalism of news records mentioning pandemics: validating a maximum entropy machine-learning model. *Journal of Clinical Epidemiology*, 75:47–55, 2016.

[21] Allan J Kimmel. Rumors and the financial marketplace. *The Journal of Behavioral Finance*, 5(3):134–141, 2004.

[22] Andrew Kohut. In changing news landscape, even television is vulnerable. *The Pew Research Center*, pages 35–38, 2012.

[23] Tuuli Koivu. Monetary policy, asset prices and consumption in china. *Economic Systems*, 36(2):307–325, 2012.

[24] A Laing. The h1n1 crisis: roles played by government communicators, the public and the media. *J Prof Commun*, 1(1):123–49, 2011.

[25] Stephan Lewandowsky, Ullrich KH Ecker, and John Cook. Beyond misinformation: Understanding and coping with the "post-truth" era. *Journal of applied research in memory and cognition*, 6(4):353–369, 2017.

[26] Agustin Maravall and Ana del Rio. Time aggregation and the hodrick-prescott filter, banco de espana, documento de trabajo no. 0108. 2001.

[27] Andrew D Oxman, Gordon H Guyatt, Deborah J Cook, Roman Jaeschke, Nancy Heddle, and Jana Keller. An index of scientific quality for health reports in the lay press. *Journal of Clinical Epidemiology*, 46(9):987–1001, 1993.

[28] Protocol. An amazing 2020 for tech stocks, in charts, 2020.

[29] Robert P Schumaker, Yulei Zhang, Chun-Neng Huang, and Hsinchun Chen. Evaluating sentiment in financial news articles. *Decision Support Systems*, 53(3):458–464, 2012.

[30] Robert J Shiller. Narrative economics. *American Economic Review*, 107(4):967–1004, 2017.

[31] Jeffrey M Wooldridge. *Introductory econometrics: A modern approach*. Nelson Education, 2016.

[32] Laijun Zhao, Hongxin Cui, Xiaoyan Qiu, Xiaoli Wang, and Jiajia Wang. Sir rumor spreading model in the new media age. *Physica A: Statistical Mechanics and its Applications*, 392(4):995–1003, 2013.