Statistical Modeling of COVID-19 Data with Stock Market Trends

Mary Stirling Brown [1] and Mingwei Sun [1]

Abstract

COVID-19 has affected every aspect of society, including the economy. One way to investigate the affect of COVID-19 on the United States' economy is through the economic indicator of the U.S. stock market. This research analyzes the effects of COVID-19, specifically the daily increases in COVID-19 cases and the vaccine percentage of the U.S. population, on different stock indexes and prices. We also analyze what sectors of the economy were most affected by COVID-19. Important results include that the stock market was not largely effected by the COVID-19 pandemic in the long-term, with supporting data ranging from June 1, 2018 to December 31, 2021. Other results include that the increasing vaccination percentage of the U.S. population also did not have a significant effect on stock prices in the long-term. A statistical linear regression model is developed to analyze a specific stock price with various COVID-19 related variables, including the daily COVID-19 case increase and the vaccine percentage of the U.S. population.

Keywords: statistical modeling, hypothesis testing, COVID-19, stock market

1 Introduction

1.1 Background

A novel coronavirus (COVID-19) originated in Wuhun, China in December of 2020 and spread around the world in a matter of a few months, leading to the declaration of the spread of COVID-19 as a pandemic. The COVID-19 pandemic in the United States has affected every

[1]Department of Mathematics and Computer Science, Samford University, Birmingham, AL USA.

aspect of people's regular lives. It has also had an affect on nearly every aspect of the government and the economy of the United States of America. One important economic indicator is the U.S. stock market, specifically the Dow Jones Industrial Average, S&P 500, and the NASDAQ Composite. Different sectors of the economy were also affected differently based on what service they offer. For example, in March 2020, the Dow Jones Industrial Average stock index decreased by 26% [5]. However, the effect of the pandemic has not been significantly studied and analyzed to see the effects on the stock prices of different companies.

1.2 Literature Review

Research done by Mazure, *et. al*, analyzes the effect of the stock market in March 2020 with the onset of COVID-19. Results found that approximately 90% of the S&P 1500 stocks had asymmetrically distributed negative returns. Some of the companies that had positive returns during this time frame were food industries, natural gas, and software technologies. Some had positive returns even up to 60%. These stocks that performed poorly were subject to high volatility during this time. Some sectors of the economy that fall under this category were entertainment, crude petroleum, hospitality, and real estate, with about 70% negative returns [5]. Current research done in this paper will expand upon this to see the long-term effect on the stock market from the COVID-19 pandemic.

Khalfaoui, *et. al*, completed research on if vaccination efforts in the U.S. had any effect on the stock market as more of the population became vaccinated. They analyzed the effect through a wavelet coherence model. The result of this was seeing a strong, significant coherence between vaccination rates and the S&P 500 index. It reflected that vaccination efforts have a positive effect on S&P 500 returns for the majority of businesses cycle frequencies. [3]. Current research done in this paper will analyze how much the increase of vaccinated people in the U.S. correlate with the stock market indexes.

1.3 Significance of Research

The research in this paper answers important questions about the economy and the pandemic. The main question it answers is if the coronavirus pandemic had a significant long-term effect on

the stock market of the United States. Questions that arise from this is what particular sector of the economy was affected the most by taking a look into specific stocks of fifteen companies. Another question that is to be answered is whether the vaccination rate in the United States had any effect on the stock market, and, if so, in what ways.

This research is significant because it can help explain and predict economic trends when disaster strikes in the United States and provide reasoning on why different companies perform better than others during these times. It also is important in that it can help better reflect the long-term effects of the pandemic on the United States economy.

The research paper is followed by four subsequent sections. Section 2 details the process of data collection and methodology of statistical analysis and interpretation. Section 3 describes the regression model used to analyze the effects of COVID-19 on the stock market. Section 4 concludes the paper by detailing some of the main takeaways, limitations, and future work of this research. The last section includes the references and resources used for this research.

2 Methodology and Data

2.1 Data Sources and Collection

A vast amount of data was collected to perform statistical analysis and modeling on COVID-19 and the stock market. The related COVID-19 data includes the daily increase of cases in the United States [1] and the fully vaccination rate [4]. Fully vaccinated means that two doses were administered for Pfizer or Moderna vaccination or one dose of the Johnson & Johnson vaccination. The daily increase of cases data collection begins on January 23, 2020, and the vaccination rate begins on January 12, 2021. The related stock data includes the three main indexes and fifteen companies across five companies gathered from Yahoo Finance [2]. Stock data for all of these begins on June 1, 2018. Data collection for all categories ends on the data December 31, 2021.

Data analysis led to the conclusion that the COVID-19 pandemic in the United States started to affect daily life in March 2020. The date March 1, 2020 became the marker to separate stock market data as before COVID-19 and after COVID-19 in this research. This is why data collection begins on June 1, 2018 to give eighteen months of data before the declared COVID-

19 marker and ends on December 31,2021 to give eighteen months of data after the declared COVID-19 marker.

The fifteen companies that are in question in this research come from the thirty companies that are included in the Dow Jones Industrial Average stock index. These companies in the Dow Jones can be divided into five companies: technology, healthcare, financial, industrial, and consumer services. The top three stocks that ranked in the Dow Jones are listed in Figure 1 with the way they were ranked in the Dow Jones at the end of 2020:

Sector	**Ranking & Stock Name**
Technology	#4: Microsoft Corp (MSFT) #8: Salesforce.Com Inc (CRM) #13: Apple Inc (AAPL)
Healthcare	#1: UnitedHealth Group Inc (UNH) #7: Amgen Inc (AMGN) #15: Johnson & Johnson (JNJ)
Finance	#2: Goldman Sachs Group Inc (GS) #6: Visa Inc (V) #12: American Express Co (AXP)
Industry	#9: Boeing Co (BA) #10: Caterpillar Inc (CAT) #11: Honeywell International Inc (HON)
Consumer Services	#3: Home Depot Inc (HD) #5: McDonald's Corp (MCD) #20: Walt Disney Co (DIS)

Figure 1: Table of top three stocks in each economic sector category

2.2 Statistical Analysis

2.2.1 Hypothesis Testing

Hypothesis testing, using the t-test, was used in this research to see how much of an effect COVID-19 had on the various stock prices. According to the Central Limit Theorem, because the sample size in this case is greater than 30, it is assumed that it follows a normal distribution. This reasoning is why the t-test was chosen over the z-test. The two sample groups being the percent increase of stock prices before and after COVID-19 took effect in the U.S. In this case, we have H_0 being the stock prices before March 1, 2020 and H_1 being the stock prices after March 1, 2020. March 1, 2020 is chosen as the dividing date since that is the time when COVID-19 became an issue in the United States. The daily percent increase of each stock is used in this

hypothesis testing. The following hypotheses are used:

$$\text{Null Hypothesis: } H_0 = H_1$$

$$\text{Alternative Hypothesis: } H_0 > H_1$$

As seen from Figure 2,the p-values vary across each stock. However, none of the p-values are less than the significant value of 0.05, so none of the null hypotheses are rejected. This concludes that the stock market was affected by other factors besides COVID-19. This information is still important because the lower the p-value, the more they were affected by COVID-19. One important one to note is that the p-value for the S&P 500 is significantly higher than the Dow Jones and NASDAQ. The Dow Jones only is based on 30 companies, which are all well-known large companies. The NASDAQ is also based on large companies, but they are focused more on technology. The S&P 500 gives the best representation of the entire market from all different sectors of the economy, which is most likely why its p-value was significantly higher than the other two indexes. This could mean that larger companies dealing with technology and industry were more effected by COVID-19 than companies from various different sectors of the market, as seen in the S&P 500.

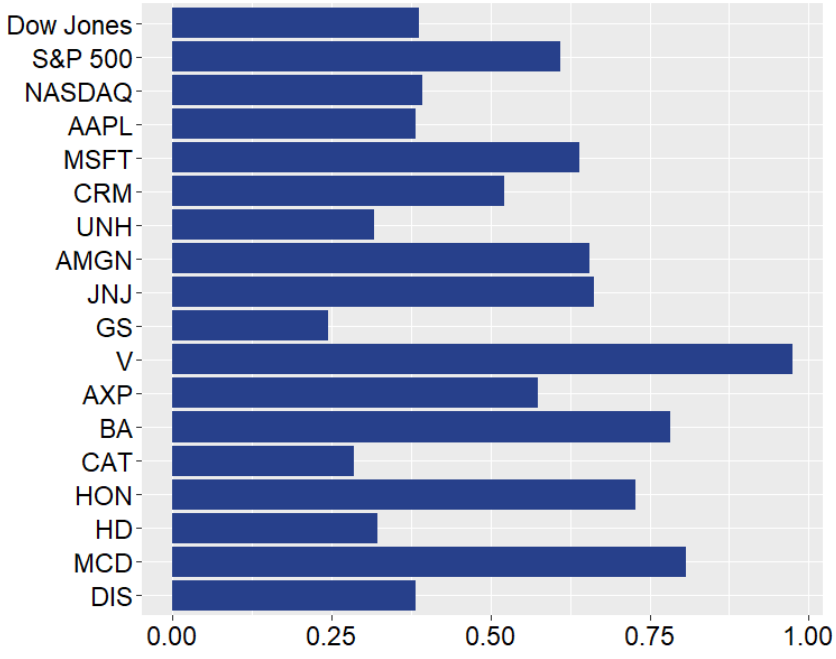


Figure 2: Bar graph of p-values for each stock price

In addition, hypothesis testing using the t-test was performed to analyze if the stock

market was affected before and after the vaccination efforts began. The vaccine primarily took effect in January 2021, so data for the hypothesis testing was separated before and after January 1, 2021. The hypothesis testing is done similar as before. However, in this case, we have H_0 being the stock prices before January 1, 2021 and H_1 being the stock prices after January 1, 2021. The daily percent increase of each stock is used in this hypothesis testing. The same following hypotheses are used:

$$\text{Null Hypothesis: } H_0 = H_1$$
$$\text{Alternative Hypothesis: } H_0 > H_1$$

As seen in Figure 3 below, the p-values are also significantly large. The confidence p-value used is 0.05. Since all the values are above 0.05, none of the null hypotheses are rejected. Since most of them are closer to 1 than 0.05, this means that the effects of the vaccine increase on the U.S. population had no long-term effect on the stock market. However, it is important to note that the Dow Jones index had the least effect from the vaccine percentage increase. This may be because it is composed of only 30, primarily large, companies, so COVID-19 may not have had as big of an effect on them as they would for smaller companies.

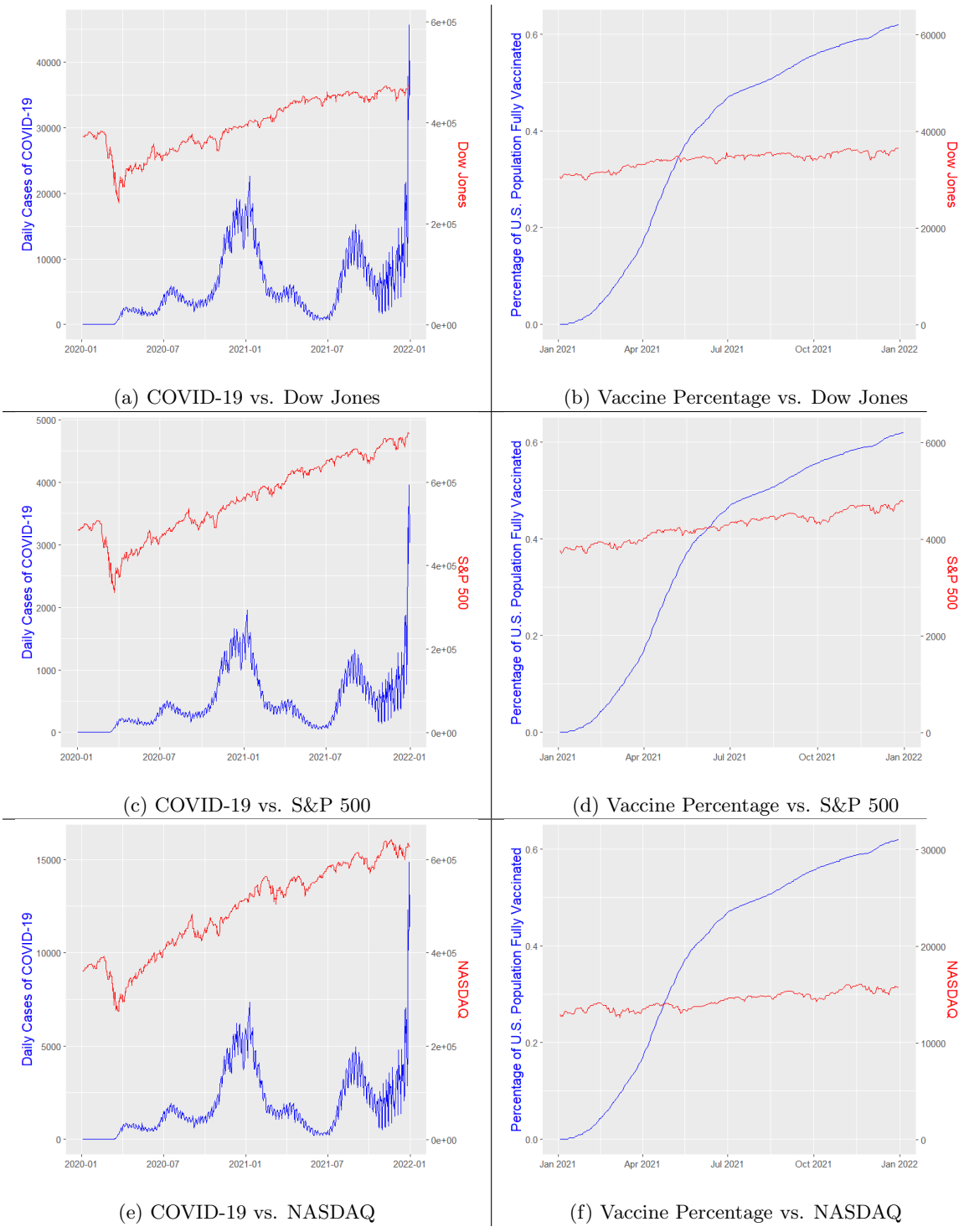
Stock Index	**P-Value**
Dow Jones	0.759
S&P 500	0.8362
NASDAQ	0.3596

Figure 3: Table of p-values for each stock index

2.3 Statistical Interpretation

3 Modeling

talk about the linear regression, the type of model, etc.



(a) COVID-19 vs. Dow Jones

(b) Vaccine Percentage vs. Dow Jones

(c) COVID-19 vs. S&P 500

(d) Vaccine Percentage vs. S&P 500

(e) COVID-19 vs. NASDAQ

(f) Vaccine Percentage vs. NASDAQ

Figure 4: Comparative graphs of COVID-19 numbers of vaccine percentage of U.S. population in relation to each stock index from January 1, 2020 to December 31, 2021.

4 Conclusion and Discussion

4.1 Summary

4.2 Limitations

4.3 Future Research

Acknowledgement

References

[1] Cdc covid data tracker.

[2] Yahoo finance - stock market live, quotes, business amp; finance news.

[3] Khalfaoui, R., Nammouri, H., Labidi, O., and Ben Jabeur, S. Is the covid-19 vaccine effective on the us financial market? *Public Health 198* (2021), 177–179.

[4] Mathieu, E. State-by-state data on covid-19 vaccinations in the united states, Jan 2021.

[5] Mazur, M., Dang, M., and Vega, M. Covid-19 and the march 2020 stock market crash. evidence from sp1500. *Finance Research Letters 38* (2021), 101690.