CSC110 Project Report
The Impact of COVID-19 on the Correlation Between Classic Economic Indicators and Monthly Median New York Stock Exchange Prices

Chenika Bukes, Krishna Cheemalapati, Aryan Goel, Dhairya Khara

Tuesday, December 14, 2021

Introduction

The COVID-19 pandemic has impacted the lives of people across the world over the past twenty-two months. Industries from energy to healthcare have undergone drastic changes because of the pandemic. The New York Stock Exchange (NYSE) has experienced unparalleled growth during the pandemic, despite many industries being negatively impacted by the pandemic, American federal debt being higher than ever, and there being abnormally high uncertainty about everything (*Bloomberg*).

The US experienced grave unemployment during the first couple months of the pandemic, yet many speculate that the outflow of stimulus checks, the less busy schedules of individuals, and the ease with which one can access the stock exchange through apps like Robinhood has led to this unexpected growth. Investors have put more money into stocks in the last 5 months than in the previous 12 years combined (*CNBC*). The average consumer spending patterns have grown to include investments in the stock market. The *S&P* 500 gained more than 16 percent in 2020, a very strong return during a year of nationwide lock downs and steep job losses in the United States (*Washington Post*).

Altogether, so far, the stock market during the pandemic can be summarised in one word: unexpected. Classic economic indicators such as unemployment rate and number of unemployed persons per job opening have been used as predicting factors for the stock exchange throughout history (Allen). However, during the pandemic, each indicator experienced extreme swings unlike anything in recent history. This led us to question the difference in the correlation between median monthly stock prices and classic economic indicators between August 2018 to December 2019 (a period of seventeen months before COVID-19 began) versus the seventeen months during the peak of COVID, from April 2020 to August 2021.

Since the pandemic could not have been predicted, we want to know whether the median monthly NYSE price action's correlation to unemployment rates and unemployed per job opening statistics was impacted by COVID-19. This project will attempt to answer the research question: **How has the COVID-19 pandemic impacted the correlation between median monthly stock prices and the classic economic indicators of Unemployment Rates and Number of Unemployed Persons per Job Opening?**

Dataset Description

Historical Stock Data

Historical stock data will be collected from Yahoo Finance using their public API. This API returns a csv file upon request. The csv file that we receive contains the "Opening", "High", "Low", "Closing, Adj.", "Close" and "Volume" of that stock for each date in the specified time period. Data will be collected from August 2018 to December 2019 for the analysis of the correlation before COVID-19 began and data will be collected from April 2020 to August 2021 for the analysis of the correlation during COVID-19.

Monthly Unemployment Statistics of the US

The monthly unemployment statistics of the US was found from the US Bureau of Labor Statistics (*"Civilian unemployment rate"*). This data is used to create a custom *.xlsx* file and stored in the local directory of the program. This file will have two sheets, with the first sheet (named "Before") representing the data before COVID-19 began (August 2018 to December 2019), and the second sheet (named "COVID") representing the data during COVID-19 (April 2020 to August 2021).

Number of Unemployed Persons per Job Opening by Month in the US

The monthly number of unemployed persons per job opening in the US was found from the US Bureau of Labor Statistics (*"Number of unemployed persons..."*). This data is used to create a custom *.xlsx* file and is stored in the local directory of the program. This file will also have two sheets, with the first sheet (named "Before") representing the data before COVID (August 2018 to December 2019), and the second sheet (named "COVID") representing the data during COVID (April 2020 to August 2021).

Computational Overview

Data Transformation and Filtering

In our *graph.py* we have 4 top-level functions which will be used in our class *Graph* method *return_info_to_graph*. These functions are:

- *get_datetime_to_closing* takes in a Pandas DataFrame object. This DataFrame object will be iterated over to create a list of tuples whose first element is the date (converted to a datetime object), and second element is the corresponding closing stock price. This method returns a list of tuples.

- *get_monthly_closing_prices* takes in a *starting_month* which is an integer between 1 and 12, and a *datetime_to_closing* which is a list of tuples. Using these variables which are passed into this method, the data is processed and returns a list of lists, where each list element contains the closing prices of the stock for each day in the month.

- *calculate_median_values* takes in a list of lists of closing prices, where each list element contains the closing prices for that stock in a specific month. The function of this method is to calculate the medians of the monthly stock prices. Using the statistics module, the median value of each list will be calculated. These median prices (stored in a new list) will then be returned and this data will act as the dependent variable on each graph.

- *get_r_squared* takes in the list of independent and dependent variables (x, y) and computes the correlation coefficient of the fitted linear model. This computation uses $numpy's\ corrcoef$ method which returns the matrix of correlations of x with x, x with y, y with x and y with y. This matrix at index[0, 1] represents the r value. We square this value to compute the correlation coefficient.

We have made four classes which all inherit from our super class *Graph*. *Graph* is an abstract class that we have made which has one private instance attribute *_ticker*. *_ticker* is a string representation of any stock ticker. We have four child classes which inherit from our abstract *Graph* class: *UnemploymentCovid*, $UnemploymentBefore$, $JobOpeningsBefore$, and *JobOpeningsCovid*.
The methods of *Graph* include:

- *__init__* is a concrete implementation which initialises $self.\_ticker = ticker$ which is an inputted stock from the user.

- *get_stock_data* is an abstract method of *Graph* that has a concrete implementation in each of the child classes. In this method, using the Yahoo Finance stock API, along with the private instance attribute *_ticker*, the stock data for *_ticker* is parsed as a Pandas DataFrame object for easier operations on the data. The difference in the concrete implementations of the child classes is the period for which the data is requested (August 2018 to December 2019 for *UnemploymentBeore* and $JobOpeningsBefore$) and (April 2020 to August 2021 for $UnemploymentCOVID$ and $JobOpeningsCOVID$).

- *get_independent_var* is an abstract method of the *Graph* class with concrete implementations in each of the child classes. The methods in each of the child classes uses pandas to read the respective excel files with data on their specific economic indicator. Using Pandas' *read_excel* method. this data will be parsed and stored in separate lists. This data will act as the independent variables for their respective graphs.
- *return_info_to_graph* has a concrete implementation in the *Graph* class. It takes in the starting month from where we want the stock data from. For example, when we call it for the 'Before' graphs, it takes in 8 and when we call it for the 'COVID' graphs, it takes in 4. The years are not required, due to the logic of the method. This method calls every method in *Graph*, and the functions *get_datetime_to_closing*, *get_monthly_closing_prices*, and *calculate_median_values*. It returns a tuple of two numpy arrays(independent variable, dependent variable).

Presentation of the Project

In $user\_interface.py$, we will use $MatPlotLib$ to present the data returned by $return\_info\_to\_graph()$ as one scatterplot graph for each class: $UnemploymentCovid$, $UnemploymentBefore$, $JobOpeningsBefore$, and $JobOpenings$ $Covid$. We will fit a linear regression model to each data set and calculate the correlation coefficient (R^2) by calling the *get_r_squared* method. This value will help quantify the correlation between the economic indicators and the chosen stock for later analysis. The R^2 value will be presented next to each graph.

All the graphs which will be plotted are Median Monthly Stock Prices vs Unemployment Rates before COVID-19, Median Monthly Stock Prices vs Unemployment Rates during COVID-19, Median Monthly Stock Prices vs Number of Unemployed Persons per Job Opening before COVID-19, and Median Monthly Stock Prices vs Number of Unemployed Persons per Job Opening during COVID-19. Their respective lines of best fit calculated by performing the aforementioned linear regression will be plotted too.

Our project will be presented in a *PySimpleGUI* interactive window. When *main.py* is run, the window will have four graphs plotting Apple's stock data as the default. The user can then search for a stock of their choice and press "Search" in the window.

Technical Requirements

- **Pandas**: This module will be used to convert our *.xlxs* files into Pandas DataFrame objects. This allows us to easily read and transform data from such files.
- **Numpy**: This module will be used to store our predictor and response variables as numpy arrays. This allows us to easily use this data to plot graphs, perform linear regression, and calculate the correlation coefficients.
- **MatPlotLib**: This module allows us to create interactive graphs quite easily by passing numpy arrays as arguments.
- **PySimpleGUI**: This module provides us with tools to present all our data in an interactive window. It allows us to easily present interactive graphs created using MatPlotLib.

Instructions Obtaining Data Sets and Running Program

1. Make sure all the python files are in the same directory.
2. Install all python libraries listed under *requirements.txt*
3. Download the excel files containing the data sets and make sure they are in the same directory.
4. Run the *main.py* file. Make sure to 'Run' the file and not 'Run file in Python Console'.
5. Recommended calls to type into the interactive search bar are $TSLA$, $AMZN$, VGT, $NFLX$, VHT, and VDE or any other stock ticker of your choice.

Changes between Proposal and Project

This project is largely based on the initial project proposal submission. After reading the TA feedback on the project proposal and experimenting with new ideas we decided to change two main things.

1. Originally, we were planning to compare three classic economic indicators and their correlations with median monthly stock prices during COVID-19. Instead, we chose to focus on investigating two of the indicators: unemployment rate and unemployment per job opening.

2. We decided that in order to truly investigate the **impact** of COVID-19 on these indicators we we will compare the relationships between these two indicators and median monthly stock prices from before COVID-19 (August 2018 to December 2019) and during COVID-19 (April 2020 to August 2021).

This means we can analyse the impact of COVID-19 on these correlations by comparison of before COVID-19 and during COVID-19 trends.

Results and Analysis

How has the COVID-19 pandemic impacted the correlation between median monthly stock prices and the classic economic indicators of Unemployment Rates and Number of Unemployed Persons per Job Opening?

To answer this research question we will investigate three stock indexes. These indexes are the Vangaurd Information Technology Index Fund ETF (VGT), Vangaurd Healthcare Index Fund ETF (VHT), and Vangaurd Energy Index Fund ETF (VDE).

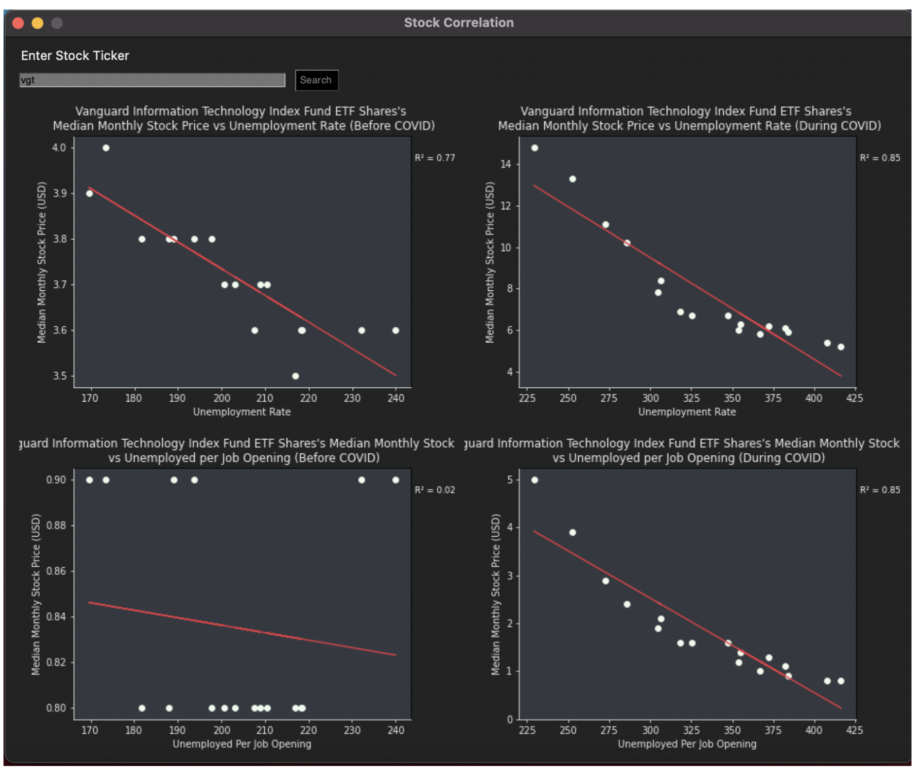

Figure 1: **The user interface of the VGT Index**

From the graphs above we can see that the unemployment rate vs the median monthly stock price before COVID-19 has a correlation coefficient of 0.77 which indicates a moderately strong positive correlation compared to 0.85 for the R^2 value of the unemployment rate during COVID-19 vs the median monthly stock price before COVID-19.

That means COVID-19 did not have a particularly meaningful impact on the strength of the correlation between technology stocks and unemployment rates. As for the correlation between unemployed per job opening vs median monthly stock prices, before COVID-19 there was a 0.02 R^2 correlation and during COVID-19 there was an R^2 value of 0.85. This means COVID-19 heavily impacted the correlation between unemployed per job opening and median monthly stock prices for technology stocks.

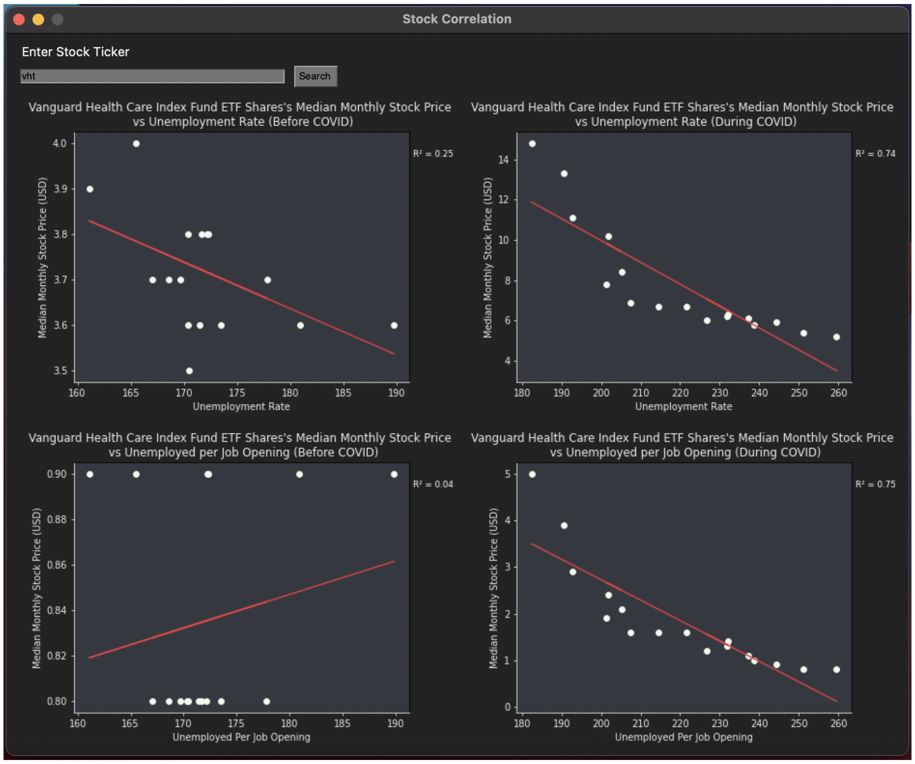

Figure 2: **The user interface of the VHT Index**

From the graphs above we can see that the unemployment rate vs the median monthly stock price before COVID-19 has a correlation coefficient of 0.25 which indicates a weak positive correlation compared to 0.74 for the R^2 value of the unemployment rate during COVID-19 vs the median monthly stock price before COVID-19. That means COVID-19 had an impact on the strength of the correlation between healthcare stocks and unemployment rates. As for the correlation between unemployed per job opening vs median monthly stock prices, before COVID-19 there was a 0.04 R^2 correlation and during COVID-19 there was 0.75 R^2 value. This means COVID-19 heavily impacted the correlation between unemployed per job opening and median monthly stock prices for healthcare stocks.

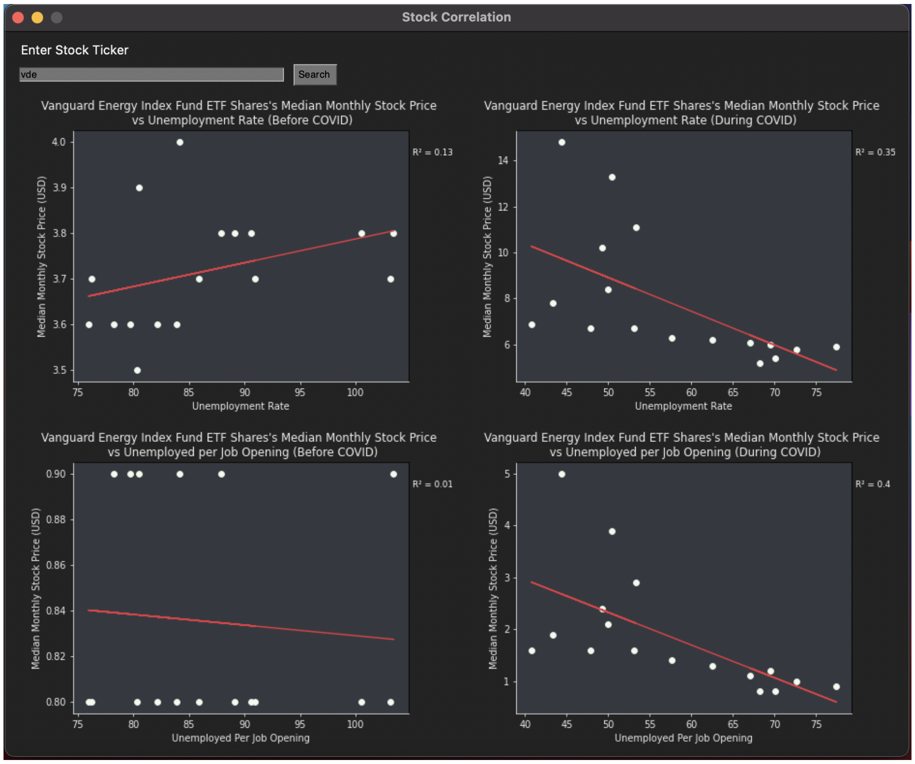

Figure 3: **The user interface of the VDE Index**

From the graphs above we can see that the unemployment rate vs the median monthly stock price before COVID-19 has a correlation coefficient of 0.13 which indicates a weak positive correlation compared to 0.35 for the R^2 value of the unemployment rate during COVID-19 vs the median monthly stock price before COVID-19. That means COVID-19 did have an impact on the strength of the correlation between energy stocks and unemployment rates. However, both of these correlations are weak and hence we can deduce from the data that in general unemployment rates does not impact energy stocks in the same way it impacts technology and healthcare stocks. As for the correlation between unemployed per job opening vs median monthly stock prices, before COVID-19 there was a 0.01 R^2 correlation and during COVID-19 there was 0.4 R^2 value. Once again, this means COVID-19 had an impact on the correlation between median monthly energy stock prices and unemployed per job openings, but the correlation is still not statistically significant.

Limitations

The primary limitation that we encountered with the data set, was that data for all of our independent variables was only available based on months, while the data with regards to the stock price is only available on a daily basis. Thus in order to solve this problem, we converted the daily stock data into monthly stock data by taking the median stock price of each month. The primary limitation with this is that the median stock price cannot accurately consider all the fluctuations that occur during a month, and cannot represent the month as a whole on an accurate basis.

Conclusions

Altogether, COVID-19 most greatly impacted the correlation between unemployed per job opening statistics and the median monthly stock price for all three indices of stocks: technology, healthcare and energy. The index which showed the least impact by COVID-19 was the VDE energy index. This might be because energy companies' stocks are more impacted by factors such as government policies and price elasticity of demand. Ultimately, there is a clear impact of COVID-19 on the correlation between unemployment rates and median monthly stock prices, as well as, unemployed per job opening and median monthly stock prices.

Further Investigation

In regards to further exploration, we could identify more classic economic indicators and not only compare the correlation between before and during COVID, but we could also compare the classic economic indicators against each other to identify which indicator correlates the best with the different categories of stocks (technology, healthcare, education etc.) before and during COVID. Furthermore, our investigation focused on the United States, using stocks only listed on the NASDAQ, therefore we could further our investigation by getting stock information from other stock exchanges to include statistics from other countries. Then we could analyse the correlation by country as well to possibly uncover other trends in the impact of COVID-19 on the correlations between stock price predictors and stock prices.

References

Allen, Bruce. “Economic Indicators That Help Predict Market Trends.” Investopedia, Investopedia, 21 Sept. 2021, https://www.investopedia.com/articles/economics/08/leading-economic-indicators.asp.

“Number of unemployed persons per job opening” U.S. Bureau of Labor Statistics, U.S. Bureau of Labor Statistics, https://www.bls.gov/charts/job-openings-and-labor-turnover/unemp-per-job-opening.htm.

“Civilian unemployment rate” U.S. Bureau of Labor Statistics, U.S. Bureau of Labor Statistics, https://www.bls.gov/charts/employment-situation/civilian-unemployment-rate.htm.

Jeff Cox. “Investors Have Put More Money into Stocks in the Last 5 Months than the Previous 12 Years Combined.” CNBC, CNBC, 9 Apr. 2021, https://www.cnbc.com/2021/04/09/investors-have-put-more-money-into-stocks-in-the-last-5-months-than-the-previous-12-years-combined.html.

“Leading Indicators - Consumer Confidence Index (CCI) - OECD Data.” TheOECD, https://data.oecd.org/leadind/consumer-confidence-index-cci.htm.

Shaban, Hamza, and Heather Long. “The Stock Market Is Ending 2020 at Record Highs, Even as the Virus Surges and Millions Go Hungry.” The Washington Post, WP Company, 31 Dec. 2020, https://www.washingtonpost.com/business/2020/12/31/stock market-record-2020/.

“Ten Ways Covid-19 Has Changed the World Economy Forever.” Bloomberg.com, Bloomberg, https://www.bloomberg.com/news/articles/2020-12-30/ten-ways-covid-19-has-changed-the-world-economy-forever.