

CSC110 Project Proposal: Effect of COVID-19 Cases on the Economy

Jacob Kolyakov, Theodore Preduta

Friday, November 5, 2021

Problem Description and Research Question

In March 2020 the S&P 500, an American index that is synonymous with its stability of growth experienced its sharpest decline in value since the housing market crash in 2008 (“S&P 500”, 2021). At the same time the United States experienced its first major wave of COVID-19 cases and was forced into lockdown. When we saw the initial sharp decline of the S&P 500 correlated with the time when COVID-19 cases first exploded in the United States, we hypothesized that there might be a connection between COVID-19 cases and the decline in value. Since March 2020, there have been several more small yet sharp declines in the S&P 500 prices as COVID-19 case numbers fluctuated in the United States, which raises the question: **To what degree do changes in COVID-19 case numbers in Canada, the United States and China correlate to changes in the economy as measured by change in stock indices?** This question arose because compared to individual stocks, which are a microcosm of the economy, indices track the performances of a collection of stocks that are considered to represent a section of the stock market/economy (“Market Indices”, 2012). The fact that the S&P500 had a sharp decline indicated to us that not only did COVID-19 affect certain stocks but the economy as a whole. We choose to explore different countries to see whether the degree of correlation is the same for different economies and what that might say about the effectiveness of countries’ strategies in handling the COVID-19 pandemic. The United States is also an interesting case to look at as they gave out relief money during the pandemic, a portion of which may have been invested into the S&P 500 and influencing its trends. This in turn, possibly, might have lessened the degree of which COVID-19 case numbers impacted the economy.

Dataset Description

This project relies on two data sources, one for financial data, and one for COVID-19 data. The COVID-19 data comes from Our World in Data which contains global COVID-19 data as a CSV, licensed under CC-BY (Ritchie et al., 2020). We keep three columns from the data, the ISO Country Code, the Date, and The number of new cases, from Canada, China, USA. An example of this data is as follows, with a larger sample in `sample-covid-data.csv`:

ISO Country Code	Date	New Cases
CAN	2020-06-25	324
USA	2021-03-19	62007

The financial data for the S&P500 is found through Yahoo Finance data, publicly available, as a CSV, through their API (“S&P 500”, 2021). The columns provided are the Date, and the Open/High/Low/Close prices. An example of the S&P500 data is as follows, with a larger sample in `sample-snp500-data.csv`.

Date	Open	High	Low	Close
2021-01-04	3764.610107421875	3769.989990234375	3662.7099609375	3700.64990234375
2021-04-19	4179.7998046875	4180.81005859375	4150.47021484375	4163.259765625

Computational Plan

We plan on analyzing both overall impact and local impact of COVID-19 case numbers on the stock market. To do this, since we are given the stock market data in absolute terms (value in USD), we must convert the data to relative daily change in the stock price. With this data we plan on comparing the two datasets in two separate ways. To identify overall trends we will calculate the correlation between the overall change in COVID-19 cases with the stock index value, for each stock index. To identify local trends we must first extract, spikes from within the data. This

can be done by looking for contiguous periods in the data where the change in index price/number of new COVID-19 cases is above a threshold. We will then try and match each spike in the index price with a spike in COVID-19 cases that occurred recently before the date. From here we can find the average market reaction time by taking the average offset between COVID-19 spikes and stock index spikes.

To accomplish this we plan on using the Pandas Python library. This library provides the critical statistics algorithms needed to compute the extent to which there is a correlation between the COVID-19 and stock index data sets. Specifically the `pandas.Series` class will be the main container of the data. It acts as a labeled array of data forming an axis on a graph. Storing both data sets as `pandas.Series` allows us to use the built in `pandas.Series.corr` function to compute the correlation constant between the two series assuming they both start at the same point and the time scale of each `pandas.Series` is the same (“pandas.Series”, 2021). `pandas.Series.corr` outputs a value between -1.0 and 1.0 where to be statistically significant the absolute value of the constant needs to be larger than 0.8 (“pandas.Series.corr”, 2021).

The program will interactively display our analysis primarily using features from the Plotly Python library, which provides graphing features. Besides what has been used in class, the main feature we will make use of is that of sliders. With sliders the user is able to vary a quantity and immediately see its result in the plots (“Sliders in Python”, 2021). We will be graphing the correlation value in respect to the delay between spikes in the COVID-19 data and the stock index data, with the local and broad data on separate graphs. The user will be able to use the slider feature to change the tolerance of what is considered a spike, thereby varying what data is filtered out, allowing them to see if (and when) there is a strong correlation between spikes in COVID-19 cases and the indexes.

References

- Market indices [<https://www.sec.gov/fast-answers/answersindiceshtm.html>]. (2012). *SEC Emblem*.
- Pandas.series [<https://pandas.pydata.org/docs/reference/api/pandas.Series.html>]. (2021). *Pandas API Reference*.
- Pandas.series.corr [<https://pandas.pydata.org/docs/reference/api/pandas.Series.corr.html#pandas.Series.corr>]. (2021). *Pandas API Reference*.
- Ritchie, H., Mathieu, E., Rod  s-Guirao, L., Appel, C., Giattino, C., Ortiz-Ospina, E., Hasell, J., Macdonald, B., Beltekian, D., & Roser, M. (2020). Coronavirus pandemic (covid-19) [<https://ourworldindata.org/coronavirus>]. *Our World in Data*.
- S&p 500 [<https://finance.yahoo.com/quote/%5EGSPC?p=GSPC&.tsrc=fin-srch>]. (2021). *Yahoo! Finance*.
- Sliders in python [<https://plotly.com/python/sliders/>]. (2021). *Plotly Python Documentation*.