CSC110 Project: Understanding the Recovery of the Economy During COVID-19

Wo Ming Boaz Cheng and Ivan Ye

Tuesday, December 14, 2021

Problem Description and Research Question

The declaration of covid-19 as a global pandemic caused a drastic market crash in the American economy during March 2020 (Walker, 2021). We are trying to figure out if the amount of deaths and positive cases during the pandemic would have any effect on a country's economy. To achieve this goal, we are going to compare the main index funds of our chosen countries, the United States of America and Hong Kong, to the amount of covid cases and deaths of the respective countries. We chose index funds to represent the economy as they are usually made up of >60% of the market capitalization. At the start of the pandemic, the tech bubble seemed to have burst, recovery seemed to be a long while away. As economic activities were halted in order to combat the pandemic, the number of positive cases and deaths were on the rise.

As the pandemic recovery progresses, so does the recovery of the economy. However, the pandemic itself is also progressing as seen by the rising amount of cases and deaths. Which leads to our research question, **'is the recovery of a country's economy affected by the number of deaths or positive cases?'** Our group was motivated to choose this research question because we are both relatively interested in trying to understand the economy more. In turn, we would then have a better understanding of the stock market. Which will let us be able to make better decisions in regards to trading during a dip, and we would also like to understand how the market reacted to one of the largest dips in recent history. With the data, we are aiming to apply computations to it to calculate its correlation coefficient and its linear regression to plot it afterwards. Doing so will allow us to have a better grasp on the effects of the pandemic in relation to the performance of the economy.

Dataset Description

covid_data_hk: a csv file from the Government of Hong Kong. It includes various covid data about Hong Kong from January 8, 2020 to the present date. The dataset has been pre-processed to only include data about the date, total number of cases as of that date, total number of deaths as of that date, and the current number of people who are in critical condition in the hospital. (Hong Kong Government, 2021)

covid_data_usa: a csv file from the Covid Tracking Project. It includes various covid data about the US from January 13, 2020 to March 7, 2021, when the project ended. Similarly to covid_data_hk, the dataset has been pre-processed to only include data about the date, total number of cases as of that date, total number of deaths as of that date, and the current number of people who are in critical condition in the hospital. (The Atlantic, 2021)

index_sp500: a csv file from Yahoo Finance. It includes various trading data about the index fund from January 13, 2020 to March 5, 2021. While it is possible to have the date be from before January, covid cases were not

tracked from back then. Since the covid_data_usa only tracks until March, we chose to stop this dataset there. The dataset has been pre-processed to only include the date about the date and the closing prices of that date. (Yahoo!, 2021)

index_hangseng: a csv file from Yahoo Finance. It includes various trading data about the index fund from January 9, 2020 to March 5, 2021. Similar to index_sp500, it is possible to have the date be from before January, but covid cases were not tracked from back then. Also, since the other index fund dataset stopped in March, we chose to do the same here. The dataset has been pre-processed to only include the date about the date and the closing prices of that date. (Yahoo!, 2021)

While most of our datasets only have data until March 2021, it should not affect our findings since the most amount of new cases and deaths happened during Winter 2020-2021. Afterwards, the impact of covid was greatly lessened, until the emergence of the delta and omicron variants. The datasets do not cover the variants, but do cover when the pandemic was the most threatening, during the time with the vaccine.

Computational Overview

First, the data is organized from the csv files to a list of their respective classes, Covid and Economy. With a list[Economy], the percent change of the closing prices of each day can be calculated with calculate_precentage_change(). Along with that, the list[Covid] is used to calculate cases per day and deaths per day with the functions cases_per_day() and deaths_per_day(). This being organizing the data and performing simpler computations on it.

With the initial computations out of the way, the points of the scatterplot can be found with the function find_all_coordinates(), which requires both percent change and cases per day. Afterwards, the line of best fit, which is also the slope, is calculated with the function linear_regression() using the formula for linear regression. This helps find a general trend line on the graph. Moreover, the correlation coefficient, which tells us the amount of correlation, is calculated through the function find_correlation_coefficient(). These two functions both use the helper function summation() which finds the sums of x, y, xy, x^2, and y^2, where x is covid cases and y is % change of the index fund. Furthermore, they are the main computations this project is built around. Since we are looking for an answer to a correlation question, the best way to view it is with a visual plot.

We are using the matplotlib to plot the graph. Since matplotlib accepts all the values of the x axis, then all the values of the y axis, the points found need to be transformed into the lists of all the points of their respective axis. The function arr_axis() converts the found points. The points for the slope line are calculated through the function slope_points(). Those points also need to be run through arr_axis().

The program then outputs two graphs with the function plot() that showcases a scatterplot with the % change of the index fund in relation to the number of covid cases per day and the number of deaths per day. A line of best fit is graphed on top of the scatterplot, which is the slope in this case of linear regression. These graphs show in a visual way if there is any correlation between our chosen variables.

Instructions

a) Install the libraries in requirements.txt
b) Download the datasets, since the datasets have been pre-processed, refer to 2). Save those datasets into the same folder as the rest of the python files.
    1) Raw Datasets:
        i) Covid data US - download found under national data

https://covidtracking.com/data/download

ii) Covid data HK - download found when scrolled to 'Latest situation of reported cases of COVID-19 in Hong Kong (English)'
https://data.gov.hk/en-data/dataset/hk-dh-chpsebcddr-novel-infectious-agent

iii) Hang Seng Index - download found under the apply button
https://finance.yahoo.com/quote/%5EHSI/history?period1=1578614400&period2=1614988800&interval=1d&filter=history&frequency=1d&includeAdjustedClose=true

iv) S&P 500 Index - download found under the apply button
https://finance.yahoo.com/quote/%5ESPX/history?period1=1578960000&period2=1614988800&interval=1d&filter=history&frequency=1d&includeAdjustedClose=true

2) Download the pre-processed datasets through UTSend:
   Claim ID: n2eJ7ptGJ9eoprog
   Claim Passcode: z7qafmawottHgKDB

c) Run main.py, two graphs should appear, one being the correlation of positive cases and the other being the correlation of deaths. To see the other two graphs, delete the commented code in main.py.

There is a zoom-in feature located at the magnifying glass at the bottom left to see the graph better, as the points are relatively spread out. The reset view is the house button located left of the magnifying glass.

Changes

Originally we were looking to answer the question **'How can the pre-pandemic and peak pandemic prices of American stocks predict the post-pandemic recovery of stocks after the COVID market crash?'** However, after working with scikit-learn a bit, we realized that our project had a scarce amount of computations within it, which is not ideal. So, we shifted the project idea to the correlation of variables, which was given from the TA feedback. Additionally, the visualization aspect moved from plotly to matplotlib.

Discussion

The discussion will be focused on the findings on the US and Hong Kong. To begin, we would like to note that the US had a relatively lenient covid policy; whereas, Hong Kong's covid policy was very strict. Also, the program currently only shows the US correlation data, to see Hong Kong's, uncomment the line in main.py.

Starting with the US' covid cases in relation to the % change of S&P 500 graph, we can see that there is only a slight correlation. The correlation coefficient located at the bottom right measures the strength of the relationship between the two variables, from a range of -1 to 1, where -1 is negative correlation, 0 is no correlation, and 1 is positive correlation. We can see that the coefficient for covid cases per day is approximately 0.04, which is very close to 0. Similarly, for the covid deaths per day, we can once again see that there is only a slight correlation of approximately 0.06 this time, which is also very close to 0. It seems that immediately, from these findings, we can conclude that there is no correlation between the two variables. Considering that with more cases or deaths, we would expect an average drop in the index. However, it also rises by a similar amount, albeit slightly less, mostly

balancing it out.

The covid cases per day in relation to the % change of Hang Seng Index graph seems to tell a similar tale of S&P 500. With a coefficient is 0.01, there also seems to be no correlation. The covid deaths per day in relation to the % change of Hang Seng Index graph is shocking, not because it is not correlated like the rest of the graphs, but because Hong Kong has never exceeded more than 5 deaths a day. In comparison to the US, where most of the points are concentrated around 1000 deaths per day. With a coefficient of approximately 0.07, it is safe to assume that the pandemic has little to no effect on the global economy.

In conclusion, our initial expectations were that there would be a negative correlation between the amount of cases/deaths and the % change of the index. Our dataset also somewhat skewed in favour of that as we used covid data from before the vaccine. However, from our findings, it can be seen that the amount of deaths or cases has no effect on the economy whatsoever since it always rises by a similar amount after a dip. This serves to conclude that the pandemic has no long-term impact on the economy, apart from the initial drop. We can also conclude that the slight market dip due to the recent omicron variant will quickly recover. Although, it will only start recovering after everything about the omicron variant is revealed and the full extent of the dip is seen.

Our research question **'is the recovery of a country's economy affected by the number of deaths or positive cases?'** was answered with a resounding no. The only limitation that we encountered was the choice of using the US covid dataset from The Covid Project. This limited our findings to only be within the original outburst of the pandemic. However, as mentioned above, this dataset was from before the time of the covid vaccine. So, theoretically it would be more accurate to investigate this dataset, since covid deaths and cases would most likely have a lesser impact on the economy post-vaccine. The next steps for further exploration comes from possibly answering why the initial fall for these index funds were so severe. Such as, knowing that an economy usually recovers, why are stocks traded for so low that the index fund loses 20-30% of its value?

References

Hong Kong Government. (2021, December 14). Data in Coronavirus Disease (COVID-19). data.gov.hk. Retrieved December 14, 2021, from https://data.gov.hk/en-data/dataset/hk-dh-chpsebcddr-novel-infectious-agent

Hunter, J. (2007). Matplotlib: A 2D graphics environment. Computing in Science & Engineering, 9(3), 90–95.

The Atlantic. (2021, March 7). Data Download. The COVID Tracking Project. Retrieved December 14, 2021, from https://covidtracking.com/data/download

Walker, A. (2021). *US Economy Accelerates as Recovery Continues.* BBC News. Retrieved November 4, 2021, from https://www.bbc.com/news/business-56932023.

Yahoo! (2021, December 14). Hang Seng Index (HSI) Historical Data. Yahoo! Finance. Retrieved December 14, 2021, from https://ca.finance.yahoo.com/quote/%5EHSI/history?p=%5EHSI

Yahoo! (2021, December 14). S&P 500 Index (SPX) Historical Data. Yahoo! Finance. Retrieved December 14, 2021, from https://finance.yahoo.com/quote/%5ESPX/history?p=%5ESPX