Predicting Economic Outlook for Post-COVID using LSTM Model

Jad Alriyabi[301360301], Insoo Rhee[301075548], Hoang Viet Truong[301385259], and Rui Yu[301321824]

Simon Fraser University, Burnaby, B.C., Canada
https://www.sfu.ca/computing.html

Abstract. This study assesses global economy by evaluates some of economic indicators and predicts the outlook with Long Short-Term Memory (LSTM) machine learning model. The data-set prepared is M2 money supply, stock prices, and the number of COVID-19 cases of eight countries. Using Jupyter Notebook and Python data analytic packages, the data-set was graphed to understand visually. Then it was evaluated in statistical summary to see the market volatility. Finally, the predictions are made using trained LSTM model.

Keywords: COVID-19 Impact on Economy · COVID-19 Cases · Stock Prices · M2 Money supply · Long Sort-Term Memory (LSTM)

1 Introduction

In light of the recent outbreak of COVID-19 around the world, economical activities decreased drastically. Airplanes were grounded, borders were closed, and some restrictions are made by the Provincial Health Officer. As a consequence, we saw the unprecedented stock market volatility all around the world. We formed a data-set by gathering three essential economic pieces; stock prices, M2 money supply and a number of COVID-19 cases of eight countries to have better understanding of global economic situation.

Once enough data-set was prepared, we wrote a program which can do three things to assess the world economic outlook. Those three things are visualization, statistical analyzing and prediction. First, we wrote a function in the program that generates a graph. The purpose of the graph is giving people an basic knowledge of time sequence of the data-set in a glance. This allows us to have a bird's-eye view in a country's economic performance during the pandemic. Second, we conducted statistical analysis and drew histograms on the change of stock prices. These calculation taught us a huge volatility in market movement which meant instability of the economy at that time. Last but not least, long short-term memory model (LTSM model) was applied to predict COVID-19 cases and stock prices. The predicted values were drawn on the graph on top of the current data-set.

2 Material

The data-set we used included COVID-19 confirmed cases by country, stock market index for each country (S&P for the U.S., S&P/TSX for Canada etc.). Money supply measured by M2 for each country. The list of countries that we support: Canada, China, France, Japan, South Korea, United Arab Emirates, United Kingdom, and United States.All the processed data can be downloaded from: https://drive.google.com/drive/folders/14pfz2bvuBpRpnjpdeSvJAdp22G2SeheM

3 Methods

The computational software we used is Jupyter Notebook which is Python IDE for data science projects. The packages we used are Matplotlib, Pandas, Numpy, Scikit-learn, and Keras. Those libraries helped us to plot, read csv data files, and apply statistical and predictive analysis.
We wrote three functions to visualize and summarize the data-set. The first function we wrote is called graph, so we can visualize the data-set in a single figure and can have intuitive understanding. To do that, we used Pandas to load the data-set, grouped them by country, and formatted their dates. Then we graphed cumulative data-set by dates in X-axis and by different scale units in Y-axis.
The second function we wrote is designed for statistics on change in stock prices. A simple statistic summary always gives us good insight of the data-set, therefore we calculated mean, median, standard deviation, skewness, ranks, and histogram on change in stock prices to see the stability of economy.

The third function we wrote does predictions using Long Sort-Term Memory (LSTM):
Developed from Recurrent Neural Network, which connects recent information to the current task, LSTM is capable of handling long-term dependencies. Typical LSTM module contains neural network layers as in Fig. 1. The Sigmoid functions (σ) are the activation functions responsible for three gates in, out, and forget. They return 0 or 1 to decide whether or not to use previous information. The tanh function pushes the all the possible values of new input between -1 and 1, then multiply with a Sigmoid function to indicate whether the new information should be updated or ignored.

C(t): Updated memory of current LSTM unit
h(t): Output of current LSTM unit
X(t): Current input

$$f_t = \sigma(W_f \cdot [h_{t-1}, X_t] + b_f) \tag{1}$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, X_t] + b_i) \tag{2}$$

$$\tilde{C}_t = tanh(W_C \cdot [h_{t-1}, X_t] + b_C) \tag{3}$$

$$C_t = f_t \times C_{t-1} + i_t \times \tilde{C}_t \tag{4}$$

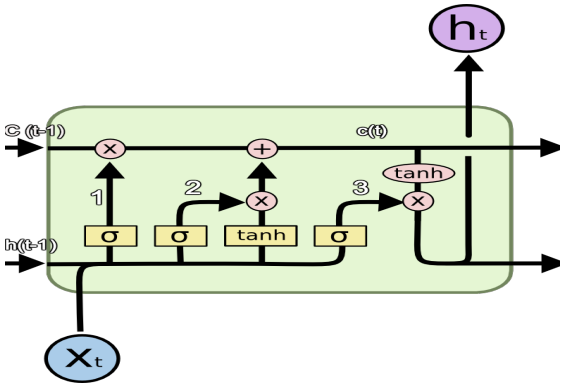

Fig. 1. LSTM Module [3]

The fundamental of LSTM is cell state, running from previous memory block (C(t-1)) to current memory block (C(t)). The forget layer decides which previous information contribute to the current state by equation (1). Then we create an update to the stored information using two network layers sigmoid and tanh as in equation (2) and (3). Finally we update the cell state using equation (5) [4].

Pseudo code for creating a LSTM model:
1. Set input_units, LSTM_units, loss and optimizer to define LSTM Network
2. Select look-back or prediction days as training window size (Number of days in the past used to predict)
3. Normalize the dataset into values between 0 and 1 (Using min_max scaler in this project) and shape it according to the training window size
4. Train LSTM Network for specified epochs and batch_size (batch mode, mini-batch mode, stochastic mode)
5. Run predictions using the trained model

In preprocessing step, data is normalized between 0 and 1, then reshaped to include data from the past 15 days (look-back = 15). This means, for instance, x_1, x_2, ..., x_{15} are used to predict x_{16}. After several trials, we decided to use only one LSTM layer with dropout 0.2 (to prevent overfitting) since it provides pretty good predictions with very limited losses. We used 80% of the data to train our model and 20% to test. After the testing stage, we forecast 30 days ahead by iteratively using the latest 15 days to predict one coming day. [5]

4 Results

The Fig. 2 below visualizes the data-set. The graph shows M2 money supply in blue, COVID-19 cases in red and stock prices in green. Every indicator went up steady though it had rough drop in March. If you are a person who follows

business section of the news, you will know what happened to asset prices during the pandemic. When COVID-19 began spreading in North America around March of 2020, stock prices tanked. At that time, FED responded by announcing unlimited QE which boosted M2 money supply and stock prices back to normal. Finally, the number of COVID-19 cases accumulated to 150 million in April 24, 2021 at the moment we are writing this report. If you are aware of the facts I listed, then it can intuitively be reminded in the Fig. 2.

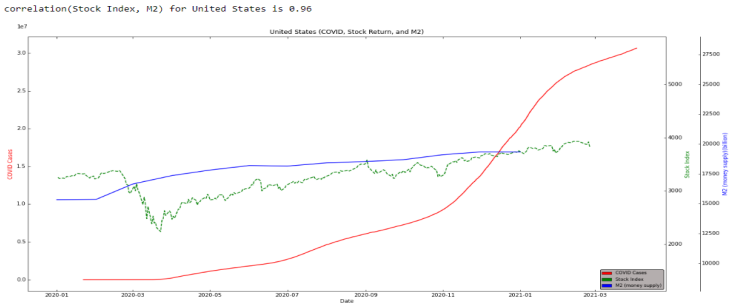

Fig. 2. United States (COVID, Stock Prices, and M2 Money Supply)

Now let us move on to the histogram of daily change in stock prices. The Fig. 3 below is the histogram and it looks like a normal distribution with a large variance and long slim tails at both sides.

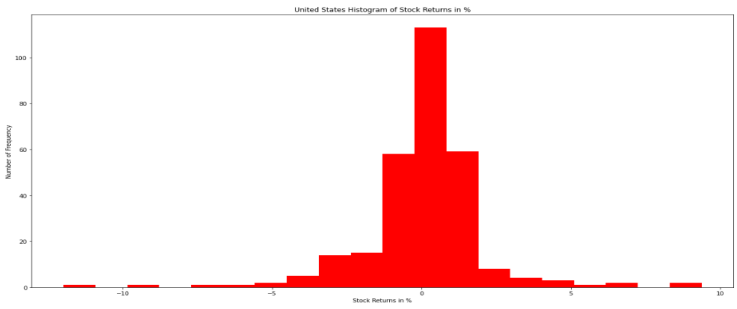

Fig. 3. United States Histogram of Changes of Stock Prices

Indeed the statics in the Fig. 4 below regurgitates the histogram. The mean is near 0.08, and the standard deviation is 2.06. We know that three sigma covers 99.7 percent of the value which is a spread between -6.18 percent to +6.18 percent change of stock prices. The median is 0.21 percent and the skewness is -0.57 which tell us more of data points are in positive area. Finally, the rank tells us that the market volatility was unprecedented in March 2020. In March

16 and 12 the United States' wealth was wiped off by -12 percent and -10 percent respectively. In March 24, and 13, it gained 9 percent each. If you know the fact that United States' wealth is around 20 trillion dollars, 10 percent of that is 2 trillion dollars. Considering that Canada's GDP is about 1.5 trillion dollars, the market had huge hiccups that happens once in a decade or perhaps only happens once in a century.

```
**United States** Statistical analysis on Stock Price daily changes.
The data's range starts from 2020-01-02 00:00:00 to 2021-02-25 00:00:00
The mean of daily change is 0.08 The median is 0.21
The standard deviation is 2.06 The skewness is -0.57
Sorted 5  of Maximum change in price
   index       Date  change
0     56 2020-03-24    9.38
1     49 2020-03-13    9.29
2     65 2020-04-06    7.03
3     58 2020-03-26    6.24
4     51 2020-03-17    6.00
Sorted 5  of Minimum change in price
     index       Date  change
285     52 2020-03-18   -5.18
286    111 2020-06-11   -5.89
287     45 2020-03-09   -7.60
288     48 2020-03-12   -9.51
289     50 2020-03-16  -11.98
```

Fig. 4. Statistical Analysis on Stock Price Daily Changes.

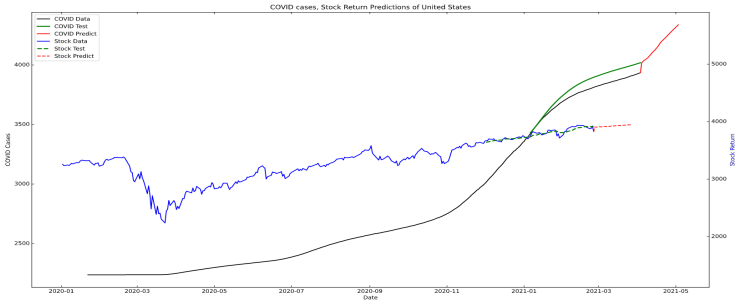

Fig. 5. United States (COVID cases and Stock Prices Predictions)

The predictions of COVID total cases and stock prices above are pretty good, the COVID model learned that the total cases should never be decreasing. Although we observe the expected increase of COVID total cases and fluctuations of US's stock prices as in Fig. 5, those values are not absolutely trustworthy. The reason is because when forecasting, instead of giving real data, the model's output is fed back as input. This creates more noise and amplifies the error, resulting in uncertain results. However, we still managed to create LSTM models predicting the real data pretty well seamlessly reasonable future data. This will set the ground base for us to apply more analysis and provide more accurate results.

5 Accomplishments and Contributions

Jad's accomplishments and contributions: I was responsible for implementing a GUI to plot our data, and analyzing government polices to help predict economic changes in our prediction model. Using tkinter in Python to create our GUI, I learned in the process that Jupyter Notebook could be a better choice for this project, due to its ability to make data analysis seamless. I was also responsible for finding the team members in the beginning and scheduling times to meet with each other to complete the task at hand.

Rui's accomplishments and contributions: At the beginning of the project, Rui created a Google Drive folder for us to synchronize works. Rui found the COVID-19 database from Our World in Data and uploaded the CSV file to the shared folder. Rui also selected 10 countries of interests. Rui selected stock index for each country and gather stock market data from Yahoo finance. Rui also gathered M2 data from each country's central bank.
Rui also contributed a lot when writing discussions part in our report.

Viet's accomplishments and contributions: The most important aspect I have learned is the LSTM model. Although I had some experiences with neural networks, I had never actually worked with a machine learning model until this project. I learned how to preprocess data and create sufficient hidden layers to produce the best predictions using Python. The difficult part of this is to understand how the LSTM model actually works, which involves understanding the Sigmoid, tanh functions and the relationships between neurons and I cannot assure I understand completely. About what I have done, I received the first code of loading data into Python from Insoo and the data from Rui, then I extended that to produce a graph of COVID cases, Stock Price and M2 % change of when given a country. Then, I also learned and implemented the LSTM model.

Insoo's accomplishments and contributions: The difficult part of working in the group was splitting the tasks evenly and each of us working on the project as owners. We are a group of four. Jad and Rui are working on major in Economics degree and minor in computer science. I am in computing science major and I also have a background in Statistics and Economics. This makes me in a weird situation because I love taking a deeper look at economic indicators and financial data with data analytic software. When we write the proposal in the beginning of the semester, we voted three to one to work on COVID-19 Economic data. I was thrilled about doing project with financial data; however, we had a bit problem in splitting the project tasks. Jad has experience in working with Python but never used Jupyter Notebook before, Rui does not have programming experience, and Viet is good at programming but he is not interested in Economics. For that reason, it was not easy to overcome and split the tasks evenly.

In the beginning when we wrote the proposal, I suggested using Jupyter Notebook for the project because it is one of the best software for data analytic. I created a Jupyter Notebook sample code and upload on Google Drive in the middle of February. I was busy with other work and I occasionally checked out Google Drive. I have not seen any follow ups from peers regarding the my sample code, so I minded my business.
Then Viet uploaded his work on graph function which worked beautifully. I modified Viet's graph function by clustering codes by indicators. I also modified it to show accumulative values in order to make the graph intuitive. Eventually I wrote a couple of lines and variables of list of countries which can work like a GUI. Later, I wrote Statistics and Histogram of changes of stock prices function to see the volatility of the stock prices. When it comes to writing the report, I wrote and modified most of the part 1 Introduction, wrote most of the part 3 Methods, and wrote most of the part 4 Results except LSTM model parts which were Viet's idea.
I also created video recording, script, and uploaded on Youtube.

6 Conclusions and Discussions

Stock Market Discussion Before further discussion in the stock market, we need to know how stock indices work. Each stock index contains stocks of different companies, at the end of a trading day, if the weighted-average price of all the stocks within a index is higher then the opening price, the index goes up. Also, in a way, the stock market reflects people's expectation of the future's economy. People would purchase stocks for capital gain if they believe the price will be higher in the future. On the contrary, if tomorrow is doom's day then everyone would want sell stocks for cash to spend today. We believe the initial market crush is a result of uncertainty and fear.

In February, the knowledge of COVID-19 was limited, people did not know the seriousness of the disease or how the government would react to the disease (full lock down). If workplaces were forced to closed down, people would need cash for their daily needs. Thus, the demand for cash suddenly raised. To cope with the needs of liquidity, individuals and companies were forced the sell their portfolios at a discount. As time passed by and people have more information related to the virus, the stock market began to rebound to normal, new COVID-cases no longer explains stock indices in a way we expect.

After doing cross-correlation for a 30 market-opening day window with a delay of 32 days between new COVID cases and S&P 500 index, we found r square to be 0.87. The market seemed to predicted the forthcoming pandemic by a month. The quick and sharp movement of the stock index demonstrated the herd effect. In February, the knowledge of COVID-19 was limited,

Raise in M2 and future concerns In response to the virus, most countries deployed Quantitative Easing (QE). Meaning the government issue currency to

buy back corporate bonds and securities from financial institutions and commercial banks. Such securities will appear on the country's central bank's balance sheet as an asset. Some argue these securities are risky and would deteriorate the country's currency credibility as the country's own currency is a liability on its balance sheet. Considering the U.S. dollars special position is global trade [1], our main focus will be on the U.S. Federal Reserve. In March 23, 2020, the Federal Reserve announced they "unlimited QE" in combination with low interest rate [2] to provide liquidity. As a result the M2 in the U.S. increased significantly, the increase in M2 from Feb 2020 to Feb 2021 is more than the increase from 2014 to 2019 combined. According to the equation of exchange, increase in money supply will eventually lead to rise in price level. Such drastic increase in M2 was the first time in the history. In response to high expected future inflation and a low interest rate environment, individuals and companies have to invest in riskier assets for higher returns. Such behaviour could also explain the booming of the U.S. stock market recently while real GDP is lower than 2019 level. Some argue the market is in a bubble right now, some even suggest another crush will happen in 2021. Only time will tell us what is going to happen next.

7 Future Work

We went over the data-set of eight different countries and the results from running our program; nonetheless, it would be interesting to have deeper analytic. We believe a single great improvement for this project is to apply analysis method and update our current model to give more reasonable predictions. We are enthusiastic to see how machine learning can make great use of past data to give highly accurate predictions.

References

1. Amadeo, K. (2020, May 22). Why the US dollar is the global currency. Retrieved April 23, 2021, https://www.thebalance.com/world-currency-3305931
2. Robb, G. (2020, March 23). Fed announces unlimited QE and sets up several new lending programs. https://www.marketwatch.com/story/fed-announces-unlimited-qe-and-sets-up-several-new-lending-programs-2020-03-23
3. Olah, C. (2015, August 27). Understanding LSTM Networks. http://colah.github.io/posts/2015-08-Understanding-LSTMs/
4. Olah, C. (2015, August 27). Understanding LSTM Networks. http://colah.github.io/posts/2015-08-Understanding-LSTMs/
5. Jose, C. (2019, September 20). Predicting Sequential Data using LSTM: An Introduction. https://towardsdatascience.com/time-series-forecasting-with-recurrent-neural-networks-74674e289816