The Relationship Between the Chinese Stock Market, the US Stock Market, and Some Other Economic Indexes Under the COVID-19 Pandemic

Jinyu Yang, Ziyi Jia and Yunxiang Mao

Cuihua Road No. 115, SSFBC, Xi'an, Shaanxi, China
E-mail: jy2080529402@163.com; 2446388808@qq.com; 1971097757@qq.com

Abstract. The COVID-19 outbreak has had a huge impact on all sectors of the world, especially the economy. Among all the economic indicators, the stock market indexes are particularly important, because the stock markets anticipate the future economic change and reflect investor confidence and thus may be the barometer of the overall economy. The Chinese and US stock markets are the most important stock markets. Because New York Stock Exchange (NYSE) and National Association of Securities Dealers Automated Quotations (NASDAQ) have the greatest market cap worldwide. And Shanghai Stock Exchange (SSE) and Hong Kong Stock Exchange (HKSE) can obviously show the conditions of economy in the mainland of China and East Asia. We collected the daily increase or decrease indexes of the U.S. and China stock markets to study their relationships and how they are affected by the pandemic. To study this problem, this paper uses three methods, including Local Similarity Analysis (LSA), Least absolute shrinkage and selection operator (LASSO) and Markov chain. We calculate the correlation between the U.S. stock indexes and China stock indexes with LSA and analyze the possible factors which lead to these correlations. LASSO is used to find out other economic indexes which affect the U.S. and China stock markets and test whether these indexes can predict the trend of future markets. Lastly, Markov chains can be used to analyze the effects of the day to the next day transmission in stock markets. We find that there is a high correlation between the U.S. and the China stock using LSA, especially, from Jan.10 to Feb.10. We find that Shanghai Stock Exchange Composite Index (SSEC) and NASDAQ are correlated to the Standard and Poor’s 500 Index (S&P500) in the US market. S&P500, confirmed cases of COVID-2019 in China (Chinese cases), USD to CNY exchange rate are correlated to the SSEC in the Chinese market. Using these indexes, the future indexes of the U.S. stock market can be predicted at a PCC correlation, the highest is 0.44 and the lowest is 0.24 for the China market. LASSO shows that the U.S. market can be more accurately predicted by the indexes. The estimated Markov models of the two stock markets are pretty similar, i.e. the probability that the value of the U.S. stocks increases two days in a row is 13.40% and the Chinese market is 11.34%. To sum up, we find that the two stock markets are highly correlated under the Covid-19 pandemic despite from some differences.

1. Introduction
Since the end of 2019, COVID-19 pandemic has been widely spreading all over the world. As of November 23, 2020, there are over 58 million cases globally. The Covid-19 pandemic affects the economic system seriously in China and the US, which means it influences the exchange rate among
these two countries, the domestic stock market indexes, and the interest rate, amid increasing number of people getting Covid-19 pandemic. Before the outbreak of novel Covid-19 pandemic pneumonia, the degree of economic market and the degree of opening up also had different effects on the stock market. Chinese investors will face the risk of government policies. Technical analysis theory is more effective in the United States than in China [1]. The conditions for wave analysis to play a role in Chinese stock market needing to be mature [1]. On the other hand, the non-market factors of Chinese stock market restrict the efficiency of the systematic market [2].

Covid-19 pandemic proliferation could devastate the economic growth. As a result, government action may not be enough to stop the stock market from falling. Many central banks have responded to the pandemic by lowering interest rates. In theory, central bank interest rate cuts should boost the economy by reducing borrowing costs and encouraging consumption [3]. So far, major central banks around the world, including the Federal Reserve and the Bank of England, have slashed interest rates. In addition, oil prices have fallen to their lowest level since June 2001, which means that investors may not be able to make relative profits [4].

In the previous literature on the stock market, there are other methods which are used to determine that the correlation between the US and Chinese stock markets and other economic indexes. For example, a large number of studies and applications of neural networks in solving business problems have proved their advantages over the classical methods without artificial intelligence [5]. In addition, the main advantage of neural network is that it can discover patterns and irregularities, and detect multidimensional nonlinear connections in data. The latter is very useful in modeling dynamic systems [6]. In other words, the subprime crisis did affect China's stock market [7]. In a sense, the sharp correction of China's stock market from February to July 2007 has greatly promoted the volatility surge in other markets. Due to the restrictions on foreign investment, the volatility of China's stock market during the subprime mortgage crisis has not been greatly affected [8].

The paper is organized by three parts. In the first part, we used LSA to study the correlation between the stock market in China and United States. In the second part, we used LASSO to study the other economic and pandemic indexes that affect the U.S. and China stock markets and whether these indexes can be used to predict the stock market value. In the third part, we used Markov chain to analyze the effects of the day to the next day transmission in the stock markets. The main goals in this paper are as follows. Firstly, we want to study we studied how the Chinese and US stock markets are correlated under the particular Covid-19 pandemic. Secondly, we analyze what factors such as interest rate and pandemic severity which would impact the trends of the stock markets in China and United State. Thirdly, we studied what the Markov properties of the two stock markets.

2. Methods

2.1. Data description

For this paper, we collect the S&P 500 index and the SSE Composite Stock Price Index, reflecting the volatility of US and Chinese stock markets. Considering other data that may affect the indexes of the stock markets of China and the US, we collect these data, including the following terms:

- Daily confirmed new cases of COVID-19 in China,
- Daily confirmed new cases of COVID-19 in the US,
- WTI crude oil,
- NASDAQ-100 index,
- Shanghai interbank offered rate,
- One-year treasury bonds in China,
- One-year treasury bonds in the US,
- CNY to USD exchange rate.

Overall, 10 kinds of data are collected, with 2 kinds of pandemic data and 8 kinds of economic data.
The range of data is excerpted from January 1st to July 6th. The data are restricted by the different time zones. The confirmed cases of COVID-19 in China and the US are recorded in the midnight in Central European Time. CNY to USD exchange rate is recorded in the midnight in Beijing Time. Other 7 kinds of stock data are recorded as the closing price of in the local time of stock exchange. For the missing data, we just delete the whole row if there is a missing data in the row. It means 10 columns of data are aligned with the uniform effective date. Next, we calculate the everyday growth of new confirmed cases of COVID-19 in China and the US. Similarly, we calculate the everyday growth rate of the 8 kinds of economic data. The reason for the different processing criteria between pandemic data and economic data is that COVID-19 may steep rise in confirmed cases. The surge of cases may affect the outcome of the analysis.

2.2. Research the relationship between the Chinese stock market and the US stock market by Local Similar Analysis (LSA)

Person correlation coefficient (PCC) is commonly used for calculating the relationships between different types of data. Suppose there are two random variables X and Y which are denoted as \( x = (x_1, x_2, \ldots, x_N) \), and \( N \) is \( y = (y_1, y_2, \ldots, y_N) \) the length of the data.

The PCC between them can be calculated using this formula:

\[
r_{xy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{(N-1)s_x s_y}
\]

In this equation (1) \( s_y \) and \( s_x \) represents the sample standard deviations of \( x \) and \( y \). Can be calculated with following formulas:

\[
s_x = \sqrt{\frac{1}{N-1} \sum (x_i - \bar{x})^2} \quad \text{and} \quad s_y = \sqrt{\frac{1}{N-1} \sum (y_i - \bar{y})^2}
\]

Local similarity analysis (LSA) is used to show the local similarity between two normal transformed sequences with same length. To deal with the data, we first rank them in ascending order. Next, we calculate the ratio of the rank divide by \( N+1 \). Then we calculate the quantiles of the ratio of the standard normal distribution. Suppose these two random variables with the same length are \( V_i \) and \( V_j \). The positive score matrix \( P_{n \times n} \) and negative score matrix \( N_{n \times n} \) are calculated by:

For, \( i, j = 1, \ldots, n \), \( P_{i,j} = 0 \), and \( S_j = \sqrt{\frac{1}{N-1} \sum (y_i - \bar{y})^2} \)

For \( i, j = 1, \ldots, n \) with \( |i - j| \leq D \), \( P_{i+1,j+1} = \max\{0, P_{i,j} + V_{i+1} \cdot V_{j+1}\} \)

and \( N_{i+1,j+1} = \max\{0, N_{i,j} - V_{i+1} \cdot V_{j+1}\} \)

and \( \text{MaxScore}(V_1, V_2) = \max(P(V_1, V_2)) \)

To get the local similarity score of two same length time series \( V1 \) and \( V2 \), LS (\( V1 \), \( V2 \)) can be calculated using MaxScore (\( V1 \), \( V2 \)) divided by the length of the time periods, as shown in following formula, \( n \) is the length of the data.

\[
\text{LS}(V_1, V_2) = \frac{\text{MaxScore}(V_1, V_2)}{n}
\]

The sign of the final local similarity should be consistent with the sign of the output of Flag(V1, V2) [9].

2.3. Measuring the correlation of one day to the next between the Chinese and the US stock market by Markov chain

For a Markov chain, the probability of transiting from a state to another state depends on the previous states. As long as every state is reachable, the Markov chain reaches equilibrium. And the number of
times visiting each state converges to some specific probability distribution, which is also called the stationary distribution [10].

In our study, we investigate how the first-day stock value influences the next-day stock value, with the data of S&P 500 index and the SSE Composite Stock Price Index. Therefore, we get three Markov chains. The first one is how the first-day china stock value influences the next-day china stock value, with the two states being decrease in the SSE Composite Stock Price Index or increase in the SSE Composite Stock Price Index. Therefore, there are four kinds of combinations of states, i.e., the transition probability1 from increase to increase, the probability1 is calculated by dividing the number of transmission of increment to increment by the number of days of increment; the transition probability2 from increase to decrease; the transition probability3 from decrease to increase; and the transition probability4 from decrease to decrease. The second one is how the first-day the US stock value influence the next-day the US stock value, included two states, the decrease in S&P 500 index, and the increase in S&P 500 index, and we got 4 kinds of probability. The third Markov chain focuses on the shifting between china stock value increase/decrease and the same/next day of US stock value increase or decrease, and there are 4 states with 8 kinds of probability. After that, we draw these three Markov chains together in 4 states.

2.4. LASSO—The factors of the Chinese and the US stock market

LASSO (least absolute shrinkage and selection operator) is a regression analysis method that not only selects variables, but also regularizes them to improve the prediction accuracy and interpretability of the statistical model [11]. LASSO solves the following problem:

$$\min_{\beta_0, \beta} \in \mathbb{R}^{p+1} = \frac{1}{2M} \sum_{i=1}^{M} (b_i - \beta_0 - a_i^T \beta)^2 + \lambda \| \beta \|$$

In this formula, bi is the object we study for, it could be S&P 500 index or the SSE Composite Stock Price Index, ai is the potential economic and pandemic indexes including the 9 collected data. The coefficient vector $\beta$ refers to the weight of each index, which is determined by lasso. The parameter $\lambda$ refers to the cross validation coefficient which is also determined by data. M refers to the total days of the data. Here we use $M=99$.

First, we figure out the other factors that may affect the US stock market. We regard the S&P 500 indexes in 99 days as object b and regard the other 9 kinds of data of all the days as factors a. We use the R package “glmnet” for our analysis. We fit the model using the basic call to glmnet and get the coefficient vector $\beta$. [12][13] Therefore, we can compare the relationship between the nonzero coefficients, the percentage of deviance, and the cross validation coefficient. After that, we use “cvfit” to find out the component that contains cross validation. "cvfit" is also an object of the glmnet package. Next step, we make the prediction for the S&P 500 in these 99 days using the data matrix based on the selected indexes output by lasso. We then compare the predicted data with the authentic data of S&P500. This process considers whether the 9 kinds of data excepted can be used to predict the future stock market values or not.

Similarly, we can take part of the days as samples instead of using the whole 99 days to make the prediction. Take the data matrix in the first 66 days for the sample and we get the predicted values in the last (99-t) days. After that, we compare the predicted values with the authentic values in the last (99-t) days. In this process, we use the past data to predict the future data. And in the prediction of 99 days, we use all the data to predict itself, and does not has anything to do with the future trend of the stock market. So the correlation between the predicted data and the authentic data in sample may be more practical than using all 99 days.

We use the same methods to compare the predicted data and the authentic data for the SSE Composite Stock Price increase rate.
3. Results

3.1. Local Similar Analysis (LSA)
LSA shows that there is a high correlation between the Chinese and the US stock markets under the COVID-19 pandemic compared to PCC.
In this section, we collect data of US stock S&P 500 and China stock the SSE Composite Stock Price Index respectively, and calculate the correlation of the increase rates between these two stocks using LSA and PCC. See Figure 1.

![Comparison between PCC and LSA result](image)

Figure 1. Comparison between PCC and LSA result

We calculate the correlation between the Chinese stock and the US stock in each time window of 30 days. Window 1 starts from Jan.1 to Jan.30. The window gap is 10 days, such as window 2 is from Jan.10 to Feb.10. According to the LSA results, the highest correlation is 0.44 and the lowest is 0.24. According to Figure 1, we can see the correlation can be reflected more apparently through LSA, and especially the highest correlation is achieved between Jan.10 to Feb.10. This time China was facing the highest level of infection by novel COVID-19. Staying indoors had closed many businesses, stores and factories, causing short-term unemployment to go bankrupt. Mass of isolation brought almost all forms of economic activity to grinding halt, except medical treatment and a few industrial. Unlike SARS 16 years ago, this epidemic outbreak coincided the golden week of the New Year holiday, greatly dampening demand during the peak season.
People stay at home and are forbidden to have parties, which greatly reduces the consumption of the masses. The development of the tertiary industry has been hit hard. The catering industry lost a lot of business. The New Year market is cold. Meanwhile, evidence shows in late January, virus already begin to spread. Therefore, time window 2 epidemic situation in USA was also serious, too. Two countries faced similar situation means it’s reasonable that COVID-19 influenced stock market.

3.2. Markov chain
Markov chain shows that the daily trend of the Chinese and the US stock market are similar in term of transition probability.
Figure 2. Markov chain (More details in the figure legend)

See Figure 2. According to the graph, it clearly shows the probability of conversion between 4 states (which four states). The probability of American S&P 500 index shift from increase to decrease is 86.60%, and the probability of China the SSE Composite Stock Price Index shift from increase to decrease is 88.66%. The probability of American S&P 500 index shift from decrease to increase is 37.11%, and the probability of China the SSE Composite Stock Price Index shift from decrease to increase is 35.05%. The probability of American S&P 500 index keep decrease is 62.89%, and the probability of China the SSE Composite Stock Price Index keep decrease is 64.95%. The probability of American S&P 500 index keep increase is 13.40%, and the probability of China the SSE Composite Stock Price Index keep decrease is 11.34%. The similar percentage values between the increase rates of the US and Chinses stock markets show the similarity of the US and China stock markets. And the stock market was going down in these periods. Because of the covid-10 pandemic, it is hard to keep increasing the stock value and it is easy to decrease the stock value, and higher percentage from increase in data to decrease in data than from decrease in data to increase.

3.3. LASSO
LASSO shows that 9 other economic indexes can actually predict the US stock market but cannot show the trend of china stock market.
See Figure 3. The graph shows the relationship between the authentic growth rate of S&P500 and the growth rate predicted by the other 9 indexes.

![Predicted and authentic growth rate of the US market based on the last 34 days](image1)

**Figure 4.** Predicted and authentic growth rate of the US market based on the last 34 days

See Figure 4. The graph shows the relationship between the authentic growth rate of S&P500 of the last 34 days and the growth rate predicted by the other 9 indexes of the first 65 days.

| Table 1. Factors influencing the U.S. stock market |
|-----------------------------------------------|
| What does the U.S. stock market affected by? |
| based on 99 days | SSEC & NASDAQ |
| based on 34 days | SSEC, NASDAQ, SHIBOR, treasury bonds & USD/CNY |

See Figure 5. The graph shows the relationship between the authentic growth rate of SSEC and the growth rate predicted by the other 9 indexes.

![Predicted and authentic growth rate of China market based on 99 days](image2)

**Figure 5.** Predicted and authentic growth rate of China market based on 99 days
Figure 6. Predicted and authentic growth rate of China market based on the last 34 days

See Figure 6. The graph shows the relationship between the authentic growth rate of SSEC of the last 34 days and the growth rate predicted by the other 9 indexes of the first 65 days.

Table 2. Factors influencing China stock market

|                         |                                                                 |
|-------------------------|-----------------------------------------------------------------|
| What does China stock market affection by? |                                                                 |
| based on 99 days        | S&P500, COVID-19 in China & USD/CNY                             |
| based on 34 days        | S&P500 & USD/CNY                                                |

Comparing Figures 4 and 6, the correlation coefficient for the US stock S&P500 is much higher than the correlation coefficient for the Chinese stock SSE. The reason for the difference is that the economic indexes collected may match with the S&P500 better, rather than SSE Composite Stock Price Index. The other reason for the strange correlation coefficient may be connected with insufficient number of days and irrelative indexes. The third reason for the low correlation of the prediction of SSEC is that there are many retail investors in Chinese stock market that caused the difficulty to predict.

4. Conclusion
The key findings of this paper include the following: We find that the correlation between China and United States is high, the high similarity may come from the following reasons: 1) Trade exchanges between China and the United States are very frequent; 2) COVID-19 affect China and the U.S. both. In addition, there is a similar trend from one day to the next for the S&P 500 indexes and the SSE Composite Stock Price Index. And the economic indexes collected can predict more objectively to the S&P 500 indexes, rather than the SSE Composite Stock Price Index. And it is more difficult to predict China stock market than the U.S. stock market.

5. Discussion
This paper also has the following limitations: First, we do not have some effective ways to get the accurate data, which means we do not have enough data because time is limited. If we have more data, we could get more accurate data during this process, 1) time range is limited 2) the types of indexes are not enough due to the limited time for this project. In addition, our data on the epidemic is also not very accurate, because there is a time delay in reporting the number of cases. Finally, we only used three methods to study the project, which would make the conclusions seem a little crude, and we should use more methods to help our research. In the future, we would like to study: 1). How the Covid-19...
pandemic affects the unemployment rate.<2). How the unemployment rate affects the GDP. 3). what actions can be done to prevent the decline of the stock.

6. Reference

[1] Jessor, R., Turbin, M. S., Costa, F. M., Dong, Q., Zhang, H., & Wang, C. (2003). Adolescent problem behavior in China and the United States: A cross-national study of psychosocial protective factors. Journal of Research on Adolescence, 13(3), 329-360.

[2] Groenewold, N. (Ed.). (2004). The Chinese stock market: efficiency, predictability, and profitability. Edward Elgar Publishing.

[3] Fawley, B. W., & Neely, C. J. (2013). Four stories of quantitative easing. Federal Reserve Bank of St. Louis Review, 95(1), 51-88.

[4] Beaudry, P., Caglayan, M., & Schiantarelli, F. (2001). Monetary instability, the predictability of prices, and the allocation of investment: An empirical investigation using UK panel data. American Economic Review, 91(3), 648-662.

[5] Zekic, M. (1998, September). Neural network applications in stock market predictions-a methodology analysis. In Proceedings of the 9th International Conference on Information and Intelligent Systems (Vol. 98, No. 1, pp. 255-263). Citeseer.

[6] Dase, R. K., & Pawar, D. D. (2010). Application of Artificial Neural Network for stock market predictions: A review of literature. International Journal of Machine Intelligence, 2(2), 14-17.

[7] Yangui, L. X. Z. (2008). A Study on Linkage between China and Foreign Stock Markets [J]. Studies of International Finance, 11.

[8] Baker, S. R., Bloom, N., Davis, S. J., Kost, K., Sammon, M., & Viratynosin, T. (2020). The unprecedented stock market reaction to COVID-19. The Review of Asset Pricing Studies.

[9] Ruan, Q., Dutta, D., Schwalbach, M. S., Steele, J. A., Fuhrman, J. A., & Sun, F. (2006). Local similarity analysis reveals unique associations among marine bacterioplankton species and environmental factors. Bioinformatics, 22(20), 2532-2538.

[10] Grinstead, C. M., & Snell, J. L. (1997). Markov chains. Introduction to probability, 405-470.

[11] Santosa, F., & Symes, W. W. (1986). Linear inversion of band-limited reflection seismograms. SIAM Journal on Scientific and Statistical Computing, 7(4), 1307-1330.

[12] Hastie, T., & Qian, J. (2014). Glmnet vignette. Retrieved June, 9(2016), 1-30.

[13] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B (Methodological), 58(1), 267-288.