Stock Price Prediction During Covid

Jiahui(Serena) Zeng - CIVE7100 Final Project

Abstract

This study investigate the relationship between stock prices and Covid data(hospitalization_count) to gain sights for the pandamic and stock pricing. Daily Sp500 stocks are used and national daily hospitalization count during 2022-1-1 to 2023-1-1 were used. Various methods such as seasonality tests(Augmented Dickey Fuller) were performed to check for seasonality of the data, Granger Causality tests were used to check for the relationship between Covid hospitalization count and stock price. Moreover, this study compares Garch(3,0)(MSE: 4.21) model with LSTM model to predict stock price during time, with result that LSTM(MSE 0.31) seems to perform better for the stock price prediction during this time period. Some limitations and future steps are discussed as well.

Introduction

This project aims to study the stock price during covid and its relationship with Covid data to gain some insights from the pandemic for both stock price prediction and future crisis purposes. Both stock price data from sp500 and Covid data from 2020-1 to 2023-1 were used. This project uses different tests to check for correlation, seasonality and compare two models (Garch model and LSTM(long short term memory) model for stock price prediction.

Background

Stock price prediction remains an intriguing subject due to its inherent volatility, which poses challenges for accurate forecasting. Extensive research (1) has underscored the role of stock prices as indicators of financial and economic conditions, emphasizing their significant impact on overall economic growth (3). The global impact of the Covid-19 pandemic from 2019 to 2023 has been profound, resulting in substantial loss of life and widespread economic ramifications.

Numerous studies have investigated the repercussions of the Covid-19 outbreak on global stock markets (2). Given the stock market's capacity to gauge the prevailing economic conditions during this tumultuous period, this project focuses on studying stock prices amid the Covid-19 crisis for specific reasons. Primarily, it aims to assess the efficacy of different models in predicting stock prices during this turbulent period. Additionally, the project aims to explore potential correlations between stock prices and pandemic-related data, such as hospitalization counts and new cases, to determine the utility of these findings for managing future crises.

Methods

This study uses the following methods:

First, Data collection: Two different types of data were gathered for this project. First, we gather the data related to Covid in this github deposit. This dataset contains 1405 points. It is the daily national hospitalization count of Covid from 2021-1 to 2023. Another dataset we used is the stock price market. This project chose sp500 daily stock price from yahoo finance ranging from 2020-1 to 2023. It specially looked into the daily close price for this project's purpose.

Secondly, Data Cleaning: We clean the data in a way such that both datasets contain the same date (ranging from 2020-1-1 to 2023-1-1). Then we use Python (google colab) to get rid of the null values in both datasets. Stock price was also transformed to returns as we perform seasonality tests to perform our models.

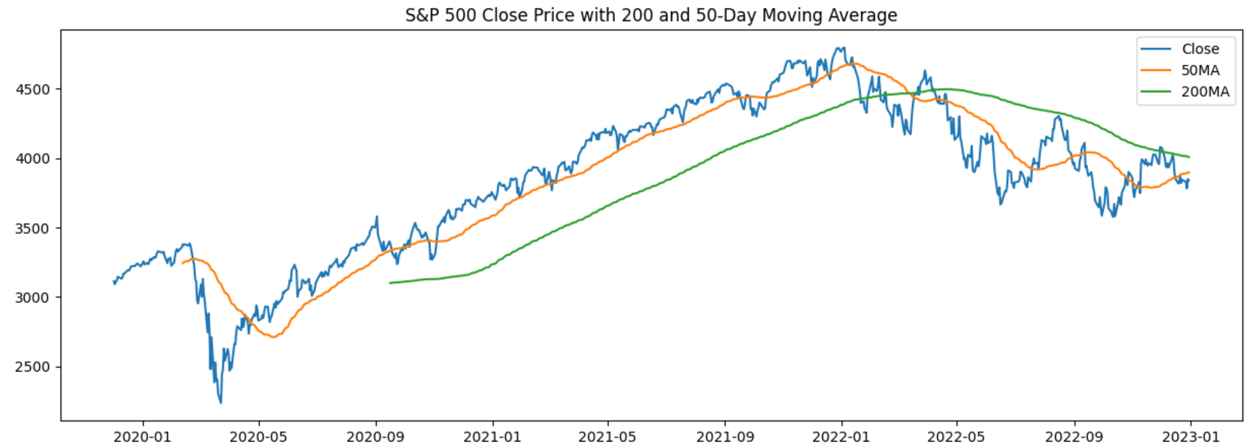
Third, Correlation test: we perform two correlation tests: one is autocorrelation and partial correlation within the stock price, and then we use Granger causality tests to compare the correlation between covid time series and our stock price.

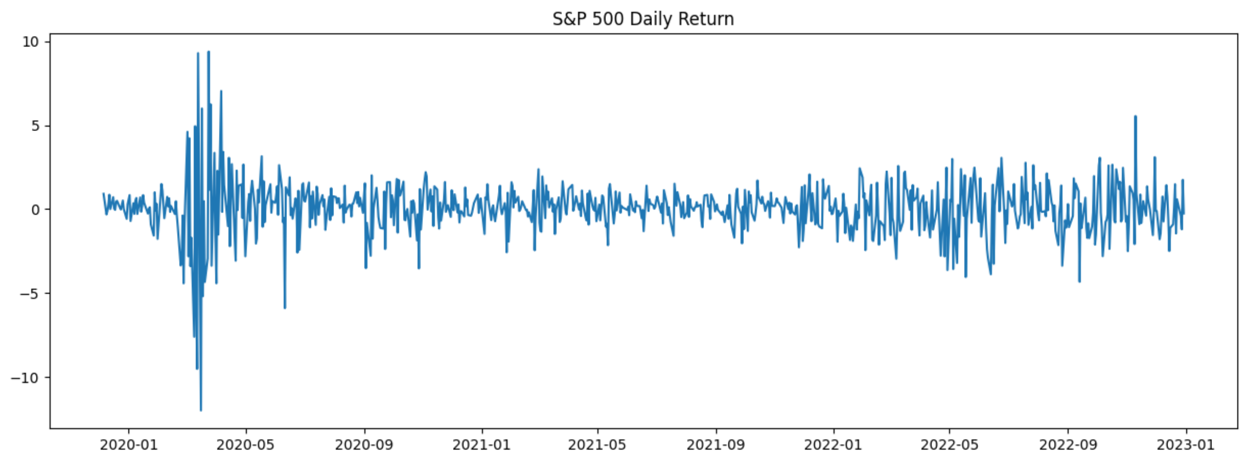
Lastly, we train two different models(ARCH-GARCH and LSTM(long short term memory model to stock prices to compare two models' performance using the test set mean squared error as the comparison.

Results

Seasonality tests:

First of all, we did the seasonality test with the stock price's close price and return price. Returns are calculated as close price's daily percentage change *100. Augmented Dickey Fuller tests are performed. Below are graphs for close stock price and returns.

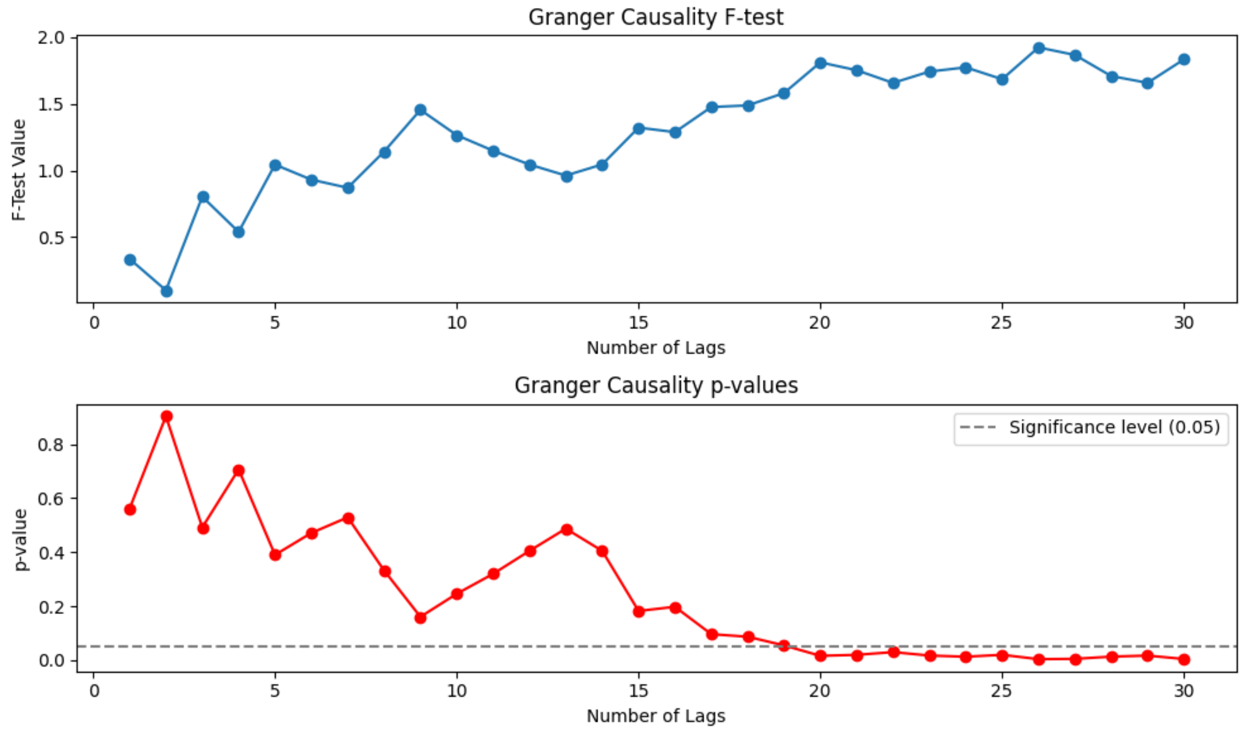




P value <0.05 is determined to check for seasonality. Result shows that close price got a p value of 0.49 while returns got a p-value of 4.55e-13. Here we can state that returns are stationary and are ok for the Arch-Garch model.

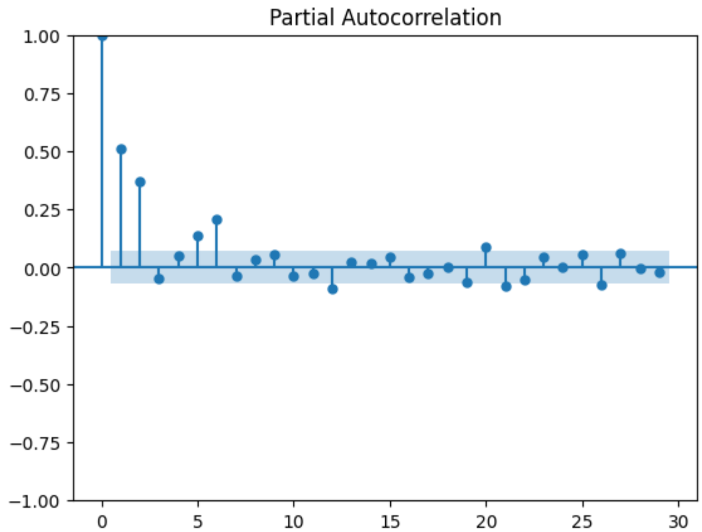
Granger tests to check for correlation between two time series:

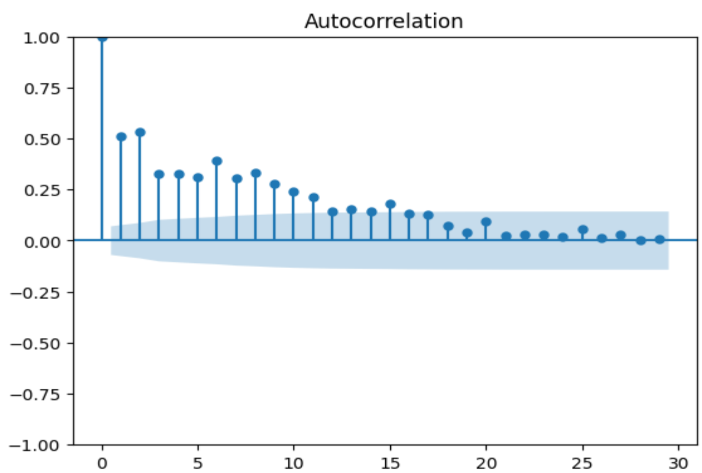
We use Granger tests to check for correlation between two time series to see if the national daily hospitality_count is correlated to daily stock price changes. When we perform granger tests for daily close price and hospitalization price, ranging from window lag of(1-30), we see no correlation as the p-value is larger than 0.05. However, we then performed a rolling moving average of windows size 30 to both the close price and the hospitalization count. We find that after lag 19, all the P-values are less than 0.05. This means that there is a significant relationship with a correlation of 1.5 to 1.75. between hospitalization count moving average and the stock closing price, but only after a lag of 19. This could indicate that the stock price is related to the hospitalization count 20 moving average before. (Shown below)



ACF AND PACF:

ACF and PACF are also performed to check for the autocorrelation of the stock price during this time period. As shown below, about 11 of the autocorrelation are significant, and about 2 of the partial autocorrelation are significant. We then use the numbers to train our model.





Models:

Garch and LSTM models are trained and tested. Both used data > 2022-6-22 as test data to calculate the mean squared error for comparison

Garch (11,2) was first performed but we find that only 3 of alpha are significant and 0 of betas are significant, so Garch(3,0) then was trained. Results are shown below. Garch(3,0) model gained a MSE of 4.42.

```
Gradient evaluations: 24
```

Constant Mean - GARCH Model Results

Dep. Variable:	Close	**R-squared:**	0.000
Mean Model:	Constant Mean	**Adj. R-squared:**	0.000
Vol Model:	GARCH	**Log-Likelihood:**	-1232.25
Distribution:	Normal	**AIC:**	2494.51
Method:	Maximum Likelihood	**BIC:**	2564.26
		No. Observations:	773
Date:	Thu, Dec 07 2023	**Df Residuals:**	772
Time:	22:47:01	**Df Model:**	1

Mean Model

	coef	**std err**	**t**	**P>|t|**	**95.0% Conf. Int.**
mu	0.0996	3.777e-02	2.637	8.366e-03	[2.557e-02, 0.174]

Volatility Model

	coef	**std err**	**t**	**P>|t|**	**95.0% Conf. Int.**
omega	0.1539	0.126	1.222	0.222	[-9.292e-02, 0.401]
alpha[1]	0.2196	0.102	2.148	3.170e-02	[1.923e-02, 0.420]
alpha[2]	0.1757	8.298e-02	2.117	3.423e-02	[1.306e-02, 0.338]
alpha[3]	0.0189	0.107	0.177	0.859	[-0.190, 0.228]
alpha[4]	0.0446	0.109	0.409	0.682	[-0.169, 0.258]
alpha[5]	1.3592e-12	8.398e-02	1.619e-11	1.000	[-0.165, 0.165]
alpha[6]	1.1267e-12	7.524e-02	1.497e-11	1.000	[-0.147, 0.147]
alpha[7]	4.5893e-12	0.115	3.974e-11	1.000	[-0.226, 0.226]
alpha[8]	4.3441e-12	5.383e-02	8.070e-11	1.000	[-0.106, 0.106]
alpha[9]	0.0136	5.624e-02	0.243	0.808	[-9.658e-02, 0.124]
alpha[10]	7.7340e-12	8.795e-02	8.793e-11	1.000	[-0.172, 0.172]
alpha[11]	0.0783	4.482e-02	1.748	8.046e-02	[-9.498e-03, 0.166]
beta[1]	0.0000	0.303	0.000	1.000	[-0.593, 0.593]
beta[2]	0.4081	0.257	1.589	0.112	[-9.540e-02, 0.912]

Covariance estimator: robust

```
Iteration:      13,   Func. Count:     88,   Neg. LLF: 1256.2631732948637
Iteration:      14,   Func. Count:     94,   Neg. LLF: 1256.2631719995102
Iteration:      15,   Func. Count:     99,   Neg. LLF: 1256.2631719994697
Optimization terminated successfully    (Exit mode 0)
            Current function value: 1256.2631719995102
            Iterations: 15
            Function evaluations: 99
            Gradient evaluations: 15
```

Constant Mean - ARCH Model Results

Dep. Variable:	Close	**R-squared:**	0.000
Mean Model:	Constant Mean	**Adj. R-squared:**	0.000
Vol Model:	ARCH	**Log-Likelihood:**	-1256.26
Distribution:	Normal	**AIC:**	2522.53
Method:	Maximum Likelihood	**BIC:**	2545.78
		No. Observations:	773
Date:	Thu, Dec 07 2023	**Df Residuals:**	772
Time:	22:50:18	**Df Model:**	1

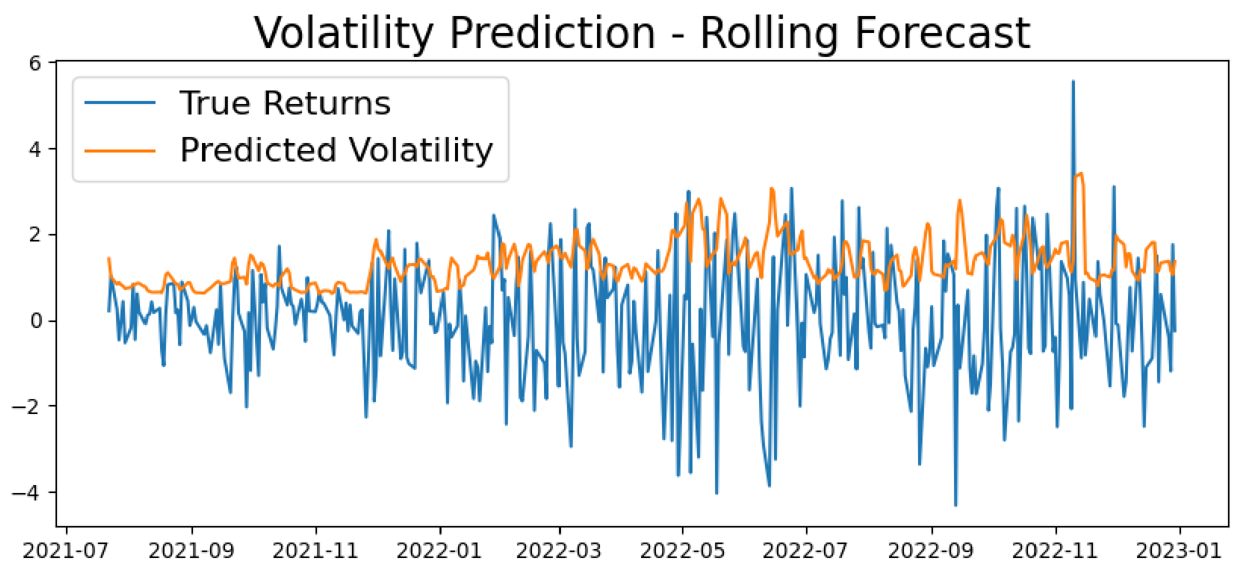
Mean Model

	coef	**std err**	**t**	**P>|t|**	**95.0% Conf. Int.**
mu	0.1052	4.001e-02	2.629	8.570e-03	[2.676e-02, 0.184]

Volatility Model

	coef	**std err**	**t**	**P>|t|**	**95.0% Conf. Int.**
omega	0.4738	0.108	4.389	1.137e-05	[0.262, 0.685]
alpha[1]	0.2802	0.108	2.587	9.694e-03	[6.788e-02, 0.493]
alpha[2]	0.3186	6.569e-02	4.850	1.234e-06	[0.190, 0.447]
alpha[3]	0.2917	8.253e-02	3.534	4.089e-04	[0.130, 0.453]

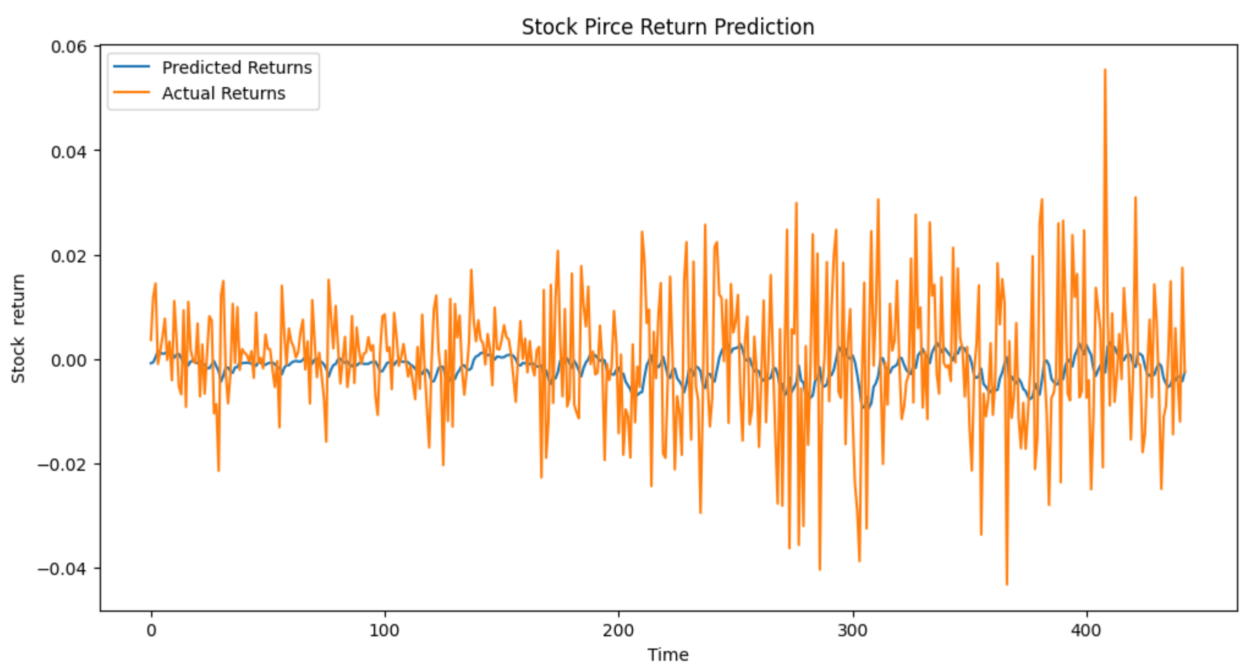
Garch(3,0)



a rolling forecast for the Garch(2,0) model

LSTM model was also trained for 30 epochs and different input length for prediction. 30 was the one with the lowest loss. LSTM uses a scalar transform to scale the data for training, and then it has layer that with LSTM layer with units 50 and a dropout of 0.2. LSTM perform with test data set with a mse of 0.31.

In this case, LSTM seems to perform better in this case for the stock price during covid.



Discussion

We can see that there is a relationship between stock price and hospitalization_count during covid. There is no correlation between those two for daily national data. However, an correlation exists for an moving average of 30. It seems like moving average of hospitalization count larger than 19 days has an correlation between 1.5-1.75 with the stock price. This could be a delayed impact of the covid to the stock price and the overall economy. Moreover, in the time of the high volatility during covid, LSTM model seems to perform better in predicting the stock price than the GARCH(3,0) model.

Limitation

There are some limitations of this project. For example, only some moving average was explored, and only some features are explored for the purpose of this project. More causality tests can be performed and more features can be used to see the data. Moreover, in the case of data training, more metrics can be trained. There could also be an overfit of the data since only data after 2022-6 was used for testing purpose. Correlation was originally not found for the hospitalization count and the stock price, after the comment from the presentation, correlation is found for the moving average of span of 30 days between hospitalization count and stock price. This also suggests that maybe other datasets other than daily dataset would give us more information for both the correlation and model comparison purpose. Future experiments with other datasets and also other features should be more impactful and should be considered. Moreover, other models that considered more features for prediction can be also be considered in the future.

Conclusion

There is an correlation between moving average(30) of Hospitalization_count during covid and that of stock price. During the time of Covid when the stock price has more volatility, LSTM seems to perform better.

Reference

1, Bai M. An Empirical Study on the Relationship between Stock Price Information and Enterprise Innovation Management Based on Information Learning Mechanism. *Comput Intell Neurosci*. 2022;2022:9425405. Published 2022 May 17. doi:10.1155/2022/9425405

2, Basuony MAK, Bouaddi M, Ali H, EmadEldeen R. The effect of COVID-19 pandemic on global stock markets: Return, volatility, and bad state probability dynamics. *J Public Aff*. Published online September 23, 2021. doi:10.1002/pa.2761

3, Chikwira C, Mohammed JI. The Impact of the Stock Market on Liquidity and Economic Growth: Evidence of Volatile Market. *Economies*. 2023; 11(6):155. https://doi.org/10.3390/economies11060155