Stock Modeling and Analysis：

Exploring and Forecasting with NVIDIA

Haonan Yang (hy2805), Yujia Bai (yb2572), Xianglong Bai (xb2166), Xinyang Chen (xc2713)

Mikhail Smirnov, Instructor

GR5010 INTRODUCTION TO MATHEMATICS OF FINANCE

I Introduction & Data Description

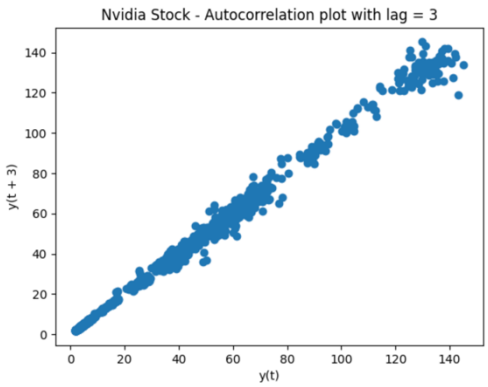
As Nvidia is a leading technology company, which has high stock performance that investors focus on. Therefore, it is very important to understand the trends that come with it and how these may be correlated with other significant market players. Our research will try to accomplish this by evaluating the performance of Nvidia stock in line with its correlations to the S&P 500 and other tech giants, such as Microsoft, in the predictive models used in trying to predict the future movements. Our goal is to analyze Nvidia's market position and uncover predictive patterns that can help investors make strategic decisions. These datasets have been compiled from sources such as the World Bank, Yahoo Finance, and FRED, covering a period given to us from the year 2009 through 2024. It includes essential financial indicators like Open Price, High Price, Low Price, Close Price, Adj Close Price, Volume of S&P 500, Google, Microsoft, Intel, and NVIDIA.

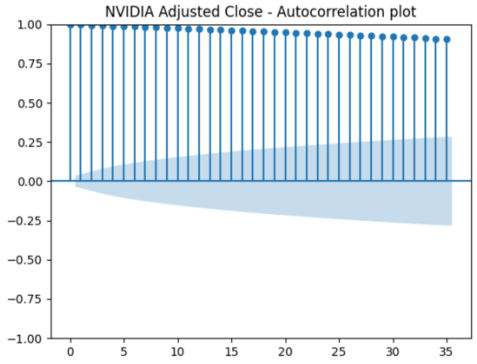
II Data Cleaning & Feature Selection

We remove all missing and infinite values in cleaning the data, and feature-scale standardization was done. We use PCA to transform the dataset into orthogonal components to maximize the variance. Lasso regression and correlation matrix was adopted within the feature selection mechanism in order to downsize less important features. Also use random forests for verification. Further, the correlation matrix shows the significant relationship between Nvidia and other technology companies and shows a strong positive correlation with Microsoft at 0.52, which means that changes in Microsoft's stock usually lead to similar movements in Nvidia's. In the effective modeling process, the features that were used include S&P 500, Google Adjusted Close, Microsoft Adjusted Close, and Intel Adjusted Close. We also calculated the 20- and 80-day moving averages of the Nvidia stock price to ensure there was no data leakage in the modeling process. In this respect, such features provide a strong ground for both the ARIMA and SARIMAX models and even the LSTM, which revealed sensitive patterns in the market behavior of Nvidia.

III Methodologies & Result

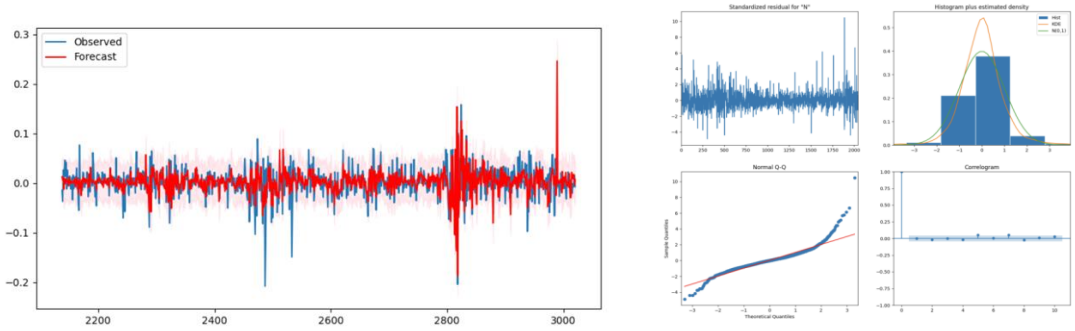
1. ARIMA Model





The three lags plot of autocorrelation shows that it is quite a stronger positive correlation, where the points almost make a straight line sloping upwards. This pattern signals that stock prices at time t are always positively related to prices three periods ahead indicate a strong linear relationship of the data points. The plot of autocorrelation with a couple of lags indicates values of the correlation of autocorrelation from -1 to 1 on different lags. There is a blue-shaded interval that shows the confidence interval within which the values of the autocorrelation are not significantly different from zero. From this plot, a high autocorrelation across the various lags is evidently seen, implying strong and persistent temporal correlations. This suggests that past price of great magnitude have an impact on future NVIDIA's stock price and the stock price of NVIDIA's stock price has a consistent trend or pattern.

2. SARIMAX Model

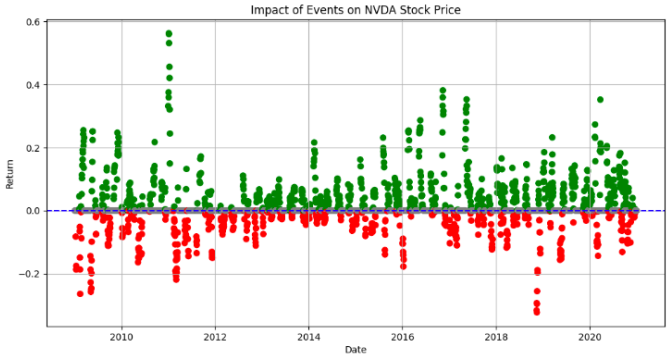


The graph of our SARIMAX model above shows a relatively good line of fit between the observed and forecasted lines of NVIDIA's stock price. From the plot above, it can be observed that our forecasted red line is following closely with the observed blue line. However, from the graph above, it can also be observed that the widths of the confidence interval are not uniform and some areas are having wider intervals. Indeed, there are a few spikes and rapid changes at certain points in both the observed and forecasted stock values, in both the mean-reverted and trended plots above, such as. This could be a moment of high volatility within the stock prices. This one brings out clearly the model's sensitivity on sudden market movements though again it re-affirms how hard and challenging it is to predict such market anomalies accurately. The residual analysis above shows that our SARIMAX model performed well though they can be improved. The residuals have no constant bias, nearly normally distributed, random residuals, and confirm that the model captures the underlying data structure effectively. However, they are not enough to allow us to make price heat predictions. So we introduce sentiment.

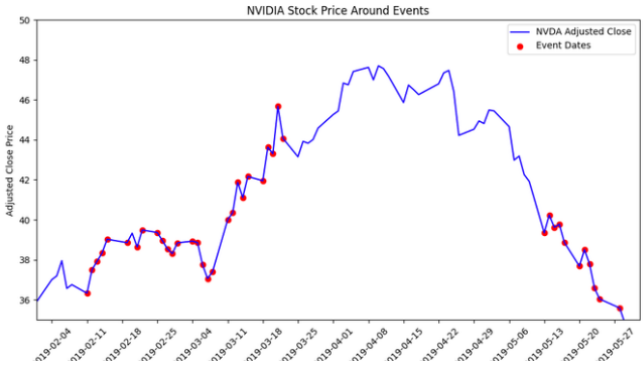
3. Sentiment Analysis

Our analysis involved the application of sentiment analysis as a tool to identify the relationship between market developments and investors' response through the analysis of NVIDIA Corporation 's stock price between 2009 and 2022. The incorporation of this approach played a critical role in the determination of the emotional and psychological motivators under market volatility. We are using the technique of news sentiment analysis to get a better prediction on the stock. (Mnih et al., 2016)

The impact of events on NVDA stock price: for analyzing the daily returns of NVDA stock with the help of a scatter plot, the provided plot notably marks the positive and negative returns. This form of color-coding enables users instantly to grasp, which periods were hallmarked by positive market sentiment, and which—by negative. In the former, the variety of green dots located above the zero return line correlates visually with positive reaction, and such reaction could have been inspired by either the positive news in the market or the firm's own success achieved through solvency, management change, corporate development, etc., in addition to positive-related manipulation. In contrast, the data points located below the zero return line and represented by red color, mark the negative periods, during which the market was pessimistic.



For the NVIDIA Stock Price Around Events, a more detailed line graph traces the adjusted close prices and overlays important market events, which are marked in red. This plot explains direct effects of given occurrences on the stock price volatility. It tells clearly how investors' sentiments relate to the actual occurrences and their subsequent influence on the market capitalization.



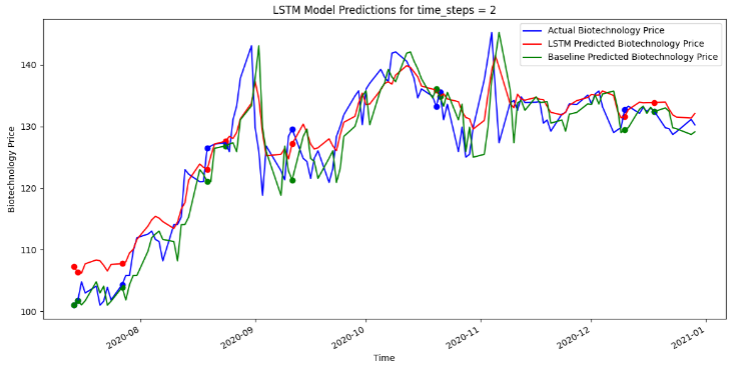
4. LSTM Model

Long Short-Term Memory networks, often referred to as LSTMs, are a distinguished variation of recurrent neural networks that excel at capturing long-distance relationships and therefore work well for sequential data such as financial time series, including exchange rate forecasting which is more accurate predictions in a complicated financial ecosystem.

Let me briefly explain the principles of LSTM. Forget gate Generating a value between 0 and 1 for each element of the cell state, with 0 meaning to ignore entirely. Input gate picks which values the network should update and creates a new candidate. Candidate values is passed through a tanh layer to create new candidate values scaled by how much the network plans to update each value. Output gate determines the next hidden, which is not included in the posted example. To implement our model to predict NVIDIA stock prices, we used LSTM networks. It is an important type of artificial intelligence that is well suited for predicting future data points given a time series.

After the feature selection and the sentiment analysis, we enter two new features, the 20-day and 80-day price averages, which allow us to have a better prediction of the trend of the stock price.

We now start the modeling and training process, our model takes the T-2 rolling window approach. To be more precise, every target day retrains the model using the data two days prior. Not only to introducing a current interpretation of events stays reliable by using the latest available data, but also avoid the data leaking. The architecture of our LSTM is specially modified for financial time series analysis. The LSTM layer is fitted for 50 units each, followed by a 0.3 dropout with 128 epochs and 128 batch size. Early Stopping is used to stop training to avoid overfitting. To prevent the data from relying on any single point, allowing the model to generalize strongly. Finally, the layer employs L2 regularization to avoid overfitting while still learning the essential long-term dependencies crucial for accurate predictions". We also implemented bidirectional LSTM layers that analyze the input sequence both forward and pitched the model for easier identification of more challenging underlying patterns. In the evaluation stage, we split the training set with 95 percent of data and rest for test. According to the performance metrics such as a low MSE, RMSE, and high R2, our model is effective in predicting the price of NVDA.



However, this is not enough to show that our model is statistically valid. So, we introduce a new t-2 baseline model. This baseline model means using the price two days ago as today's price since there is a strong correlation between the prices of stocks. Visual analysis was crucial in analyzing the model. Plots were used to visualize the

predictions with red dots depicting the predicted stock and the blue line representing the original stock price. For example, on 17th July 2020, the NVDA stock price was 102.986656, and our LSTM prediction was 103.39015, and baseline shows 104.803894 suggesting our LSTM model for NVDA price prediction was a true representation of the stock price. Also for the result, the Average MAE between LSTM predictions and actual values is 2.422473771920197, which is better than Average MAE between baseline T-2 predictions and actual values 3.437822635593221. Our LSTM model has a smaller mean absolute error compared to the baseline.

IV Conclusion

To make the dataset suitable for analysis, from data cleaning and feature selection, where PCA and Lasso regression are used at the first steps of our research to formulate the base we had picked the best, but still imperfect, models to describe the performance of Nvidia's stock — ARIMA and SARIMAX. Both models gave many core insights on how the time increase for the Nvidia's stock affects it but did not work as precise prediction tools when the task is to predict the future based on the past data. Hence, next, we conducted a sentiment analysis, which helped us to refine our dataset and open some additional insights – on how the results and the stock itself is affected by the general market sentiment and specific temporal events. With the knowledge gathered from this analysis, we designed the LSTM model, which has proven itself to be particularly useful when predicting financial time series due to its ability to get integrated long-term trends and dependencies. Indeed, this model showed good performance in predicting Nvidia's stock. Thus, the continuous methodology enhancement showed such results: pulse field. We still need to keep testing real-time data as well as that sentiment analysis to improve the Accuracy of the model

Reference

Xiao, Q., & Ihnaini, B. (2023, March 20). *Stock trend prediction using sentiment analysis*. PeerJ. Computer science. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10403218/

Yahoo Fianance
https://finance.yahoo.com/quote/NVDA/news/

Mnih, V., Badia, A. P., Mirza, M., Graves, A., Lillicrap, T., Harley, T., ... & Kavukcuoglu, K. (2016). Asynchronous Methods for Deep Reinforcement Learning. arXiv. https://arxiv.org/abs/1607.01958