**MSc Project - Reflective Essay**

|  |  |
| --- | --- |
| **Project Title:** | Comparison of stock market performance among the US and Chinese tech companies, with respect to the Impact of the Covid-19 Pandemic |
| **Student Name:** | **Wong Chuen Lik Daniel** |
| **Student Number:** | **210763720** |
| **Supervisor Name:** | William Marsh |
| **Programme of Study:** | MSc in Computing and Information System |

**Introduction**

This project grew out along with the current uncertain environment across the world, including the impact of the Covid-19 pandemic, the abnormal inflation environment, the ongoing war between Russia and Ukraine, and the regulatory scrutiny about anti-monopoly in both the US and China. Since Dec 2019, the Covid-19 pandemic has brought an unprecedented impact on society and our daily lives. This pandemic has resulted in more than 500 million confirmed cases and more than 6,345,500 fatalities worldwide, with the rising new Covid-19 variant (World Health Organization, 2022). While different businesses and industries have paid a lot of effort to adapt into this pandemic, the US government - Federal Reserve announced an imperative plan in Nov 2021 to start tapering and raise the interest rate due to the severe inflation environment, which is known as a ‘hard landing’ action. In June 2022, the Consumer Price Index (CPI) even reached a historic high of 9.1%, which was the greatest increase since November 1981 (U.S. Bureau of Labor Statistics, 2022). It revealed the necessity of the US government’s monetary measures and thus we could foresee an increasing instability in the financial market.

Due to the scope of our project, we did not cover the impacts of the Russo-Ukrainian War and the anti-monopoly policy against tech giants in the US and China in our report, which have actually significantly impacted the stock market. For instance, the geopolitical conflicts in Russia’s invasion have led to a high commodity price, where we introduced that there are complicated Intermarket relationships behind the primary markets. Meanwhile, in 2021, Joe Biden signed a new executive order about anti-competitive practices in America’s Big tech sector, where similar legislations about anti-monopoly were introduced by the Chinese government in the same year (CNBC, 2022). From the investor’s perspective, those regulatory pressure have posed considerable risks to the tech businesses and thus caused a consequential fluctuation in the stock market in recent years. Hence, the combination of these factors contributed to my interest of this project.

**Strengths and Weaknesses**

This report presents 3 main parts which are strongly intercorrelated. These are literature reviews of the current macroeconomic situation, exploratory analysis of the stock market performance in the US and China tech industries, and the data modelling of stock price prediction.

Different considerations have been taken into account at the beginning of our work. For example, when we try to analyse when will the peak of stock price cease among tech companies, it is necessary to understand what are the major factors to motivate the fluctuation of stock market prices. Hence, prior to the stock price forecasting, we attempted to explain and analyse the causal relationships between the stock performance and the external factors in literature reviews and exploratory analysis. With the aid of data visualization, it is convenient to spot out the differences between different attributes in the different time periods.

Notwithstanding, it is indeed impractical to compare the US stock market performance with those in China on the same scale and criteria, given that different country has their own economic and political conditions. As the paying ability of tradable goods in China is much lower than those in the US, it creates a comparative advantage for Chinese businesses as it can use a lower opportunity cost to produce the same good or service than its trading competitors. Besides, there are many political factors involved in the development of Chinese tech companies. The People’s Republic of China, also known as China, has had a centralized and top-down system. In spite of the Chinese economic reform since 1978, the Chinese government is still having a high degree of control over society including the tech companies. Despite both US and China have implemented similar policies regarding anti-monopoly, it is deemed that the Chinese government has a more stringent policy and all of these contributed to the result of financial recession in China market currently. Hence, this makes the comparison of market performance between the US and China tech companies very challenging, given that the Chinese tech companies are rapidly expanding and integrated globally.

Apparently, it is impossible to cover all considerations and factors beyond the stock market volatility in this report, as there are too many factors associated with the market and each of them can act as an independent topic of study. For instance, apart from the macroeconomic factors, investors often make use of different technical indicators to determine the market signal and thus predict the stock prices, such as Moving Average Convergence Divergence (MACD), Relative Strength Index (RSI), Stochastics (KDJ), etc. Most of them are so-called momentum indicators that are commonly used in technical analysis and quantitative trading, and these measures are usually deemed as more reliable than macroeconomics factors in investment decision-making, given that news and media are sometimes delayed or even manipulated by government or institutions.

In the data modelling section, two segments of the supervised learning algorithm, that are classical learning models (Polynomial Regression & SVR) and time series models (ARIMA model) have been chosen for implementation and evaluation. As we proposed the significance of macroeconomic factors in the stock market performance, we intended to utilize these macroeconomic variables in our training models and compare their effectiveness with the ultimately proved model, that is ARIMA model. ARIMA model was chosen as it is a popular model in short-horizon forecasting and it is widely adopted by the financial industry as “the ‘standard statistical instrument for forecasting asset returns” (Huijian et al., 2020). All in all, it was proved that the integration of macroeconomic variables could actually enhance the model performance of classical learning models. Hence, the aims of this project have been achieved.

In the meantime, two major weaknesses are identified in the data modelling section. That is, the untransparent process of algorithm based on built-in libraries and the difficulty of data modelling on the actual stock market. The details are described as follows.

**Untransparent process of machine learning algorithms (black box)**

It is always difficult to understand and extract all those mathematical theorems beyond the machine learning models in the coding process. Even though we can call out the built-in functions to allow the machines to perform incredible machine learning tasks, we actually don’t know the inner workings of the algorithm. For example, when we were using polynomial regression to predict the stock price, we might be able to obtain some information about the generic algorithm on the official website (sklearn library) or other academic articles, yet we did not know how the machine perform the task in details, such as the steps that have taken through or the respective weighting of each variable.

**Computationally resource**

Despite the incredible result in the ARIMA model, it is often computationally expensive to perform machine learning tasks. ARIMA essentially is an integration of two models – the autoregressive model and the moving-average model. During the process of forecasting, it is predicting every data point based on the errors made by previous data points, and thus it is required to initialise and train the model repeatedly in an iterative loop given the length of the dataset. Furthermore, in order to find out the optimal parameters (p,d,q) manually, we need to observe the statistical matrix by repeating the modelling process multiple times. Hence, the long-running time makes this project more time-consuming and infeasible in real-life prediction.

**Obstacles to long-term forecasts**

Given that the ARIMA model is proven good at the short-term forecast, both the classical learning models and the ARIMA model are not ideal for long horizon forecast, especially in time series modelling. In fact, neither a model in theory nor reality has the ability to predict the actual stock price accurately in long term. In our modelling, there is still a long way to achieve a long horizon forecast, and apparently, we need more data and a more advanced technique to do so.

**The difficulty of modelling in the actual market**

As discussed above, it is extremely difficult to create a reliable machine learning model given the current unstable market. While we were attempting to predict the stock price based on the independent variable (EMA\_10 or the actual price), the movement of stock market price was also affected by the algorithmic trading and stock analysts’ decision-making concurrently. In other words, the independent attribute of a stock price is actually autocorrelated with the results of numerous machine learning models in the financial market and investment banking. On one hand, this violates the fundamental assumption in time series modelling, where it requires stationary data as an input attribute without autocorrelation in data. On another hand, this makes the stock prediction much more challenging, as the machine learning technique is difficult to predict the psychological factors of humans, particularly in the recent abnormal trading behaviour (i.e. GME/ meme stocks).

**Future work**

In future research, more economic attributes can be added as predicting attributes in the data modelling process. On the other hand, more advanced learning techniques, such as SARIMAX, deep learning, neural network, and XGBoost, can be further developed in stock price prediction.

For example, one of the potential independent attributes is known as the “negative profit margin” of a specific company. Simply say, the stock price is highly associated with the prediction of the company’s profitability in the future. Therefore, it is expected that a company with a negative net profit margin would have relative more pessimistic prospects made by stock analysts. A negative profit margin means when a company’s expenses are larger than its revenues for a specific period. Although many companies have negative profit margins due to external factors, such as the company’s rapid growth pace and heavily investing in its technology/clients, negative net income is deemed as a warning signal as it implies considerable risk or an unsustainable business model in the company. Hence, a negative profit margin could be used as an independent feature to determine the stock price in our data modelling section.

Regarding the data modelling section of using ARIMA in our report, indeed there are many variations among ARIMA-based models, such as ARMA, SARIMA, ARIMAX, and SARIMAX models. All these models can achieve decent scores on most time-series problems with well-tuning parameters. Among all, SARIMAX could be used in our future work as it appears as the most comprehensive model which can include seasonality and exogenous variables in the model training. In other words, this model takes into account external data, which could be the macroeconomic factors (i.e., Covid-19 new cases, CPI, Unemployment rate, and OxCGRT index) that have been used in our classical learning model section.

Furthermore, there are numerous advanced machine learning models that are available to enhance the computing power of forecasting. It includes neural networks, deep learning, and XGBoost, where the contexts and difficulties are in ascending order. Neural networks are a subset of machine learning and essentially are the fundamental element of deep learning algorithms. A simple relationship between Artificial intelligence, machine learning, neural networks, and deep learning is shown below. In which, artificial neural networks (ANNs) are comprised of numerous nodes, where each node is intercorrelated and has an associated weight and threshold. If the output of any individual node is above the specified threshold value, the node will be activated and send data to the next layer of the network (IBM, 2020). For instance, in stock price prediction, when an abnormal trading volume and momentum is observed, the neural network can make use of external factors such as the content of social media as the predicting attributes in data modelling for a better prediction. In fact, it was already proved that a model that incorporates social media text features which were derived from social media based on deep learning technology can better predict stock prices (Xuan, Jiachen and Zhijun, 2021). Hence, we could also make use of online news and comments, which often reflect investors’ emotions and attitudes toward stocks, to predict the stock prices in future work.

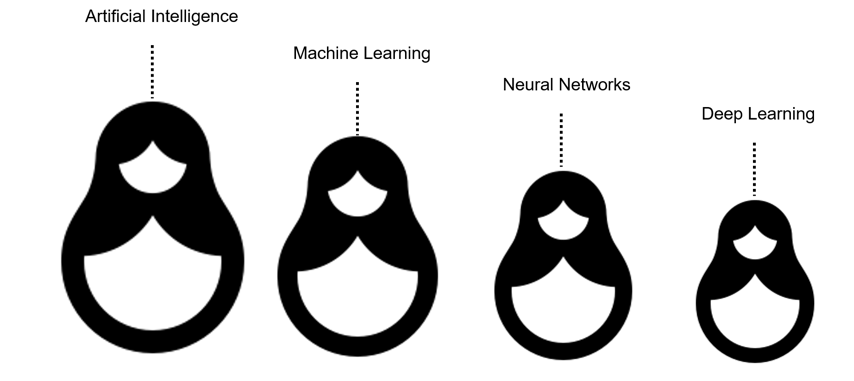


Figure 1: A graph showing relationship between Artificial intelligence, machine learning, neural networks, and deep learning

**Relationship between theory and practical work produced**

In the exploratory analysis section, we studied that both Netflix and Zoom have exploited and enjoyed the advantages brought by the Covid-19 pandemic. It was anticipated that Netflix might have a growing stock price above than average or similar to the trend in Zoom. However, the actual stock price of Netflix did not perform as good as we expected in theory, with a maximum increase of 200% in Nov 2021 and started deeply declining since 2022. Upon studying the financial news and articles, we found that the slump in Netflix’s stock price may be due to the keen competition in the streaming market and the signal of saturated growth of Netflix’s subscribers. But most importantly, we learnt that the stock price is not solely influenced by macroeconomic factors, but also the market emotion and the prospects made by stock analysts.

In the data modelling section, we proposed that macroeconomic fundamentals play a significant role in stock price prediction. However, it is indeed quite intuitive to choose the statistic number of Covid-19 new cases as our predicting attribute. There is no causal relationship between the change in covid-19 new cases and the fluctuation of stock price, yet we can make use of the covid-19 statistical data to reveal the severity of the pandemic. In comparison, a number of studies have considered specific news variables such as economic news and health news in predicting stock returns (Calomiris & Mamaysky, 2018). They proposed that both positive and negative news content could affect the investment decisions made by individual investors. Obviously, it makes more sense in theory given that it usually needs more advanced techniques and effort to do so.

Lastly, the modelling performance in Polynomial Regression differed from that in the Support Vector Regression (SVR) model. Both models have utilized the same set of predicting attributes in data modelling, yet our polynomial regression had a worse result than the SVR model. In theory, polynomial regression is a variation of the linear regression model which could capture multiple independent attributes in an nth degree polynomial equation to establish a ‘curved’ line. To compare, SVR is an integration of the Support Vector Machine (SVM) and Regression model, where SVM is widely used in classification tasks as it produces high accuracy with less computation power. Along with the development of machine learning models, SVM might even have a better predictive result than neural networks on a small dataset. Apparently, both polynomial regression and SVR have their strengths and weaknesses, but the power of SVR/SVM seems to outperform polynomial regression on this occasion.

**Legal, Social Ethical Issues, and Sustainability**

The major legal and ethical consideration of this project is the manipulation of data sources. All data sources are retrieved from the official website of authorised institutions. No data were modified without a legitimate or reasonable purpose. Also, there are no obvious sustainability issues with this project, as all codes and software are open-sourced.

References:

World Health Organization (2022). WHO Coronavirus (COVID-19) Dashboard. Available at: <https://covid19.who.int/> (Accessed: 7 Aug 2022)

U.S. Bureau of Labor Statistics (2022). Consumer prices up 9.1 percent over the year ended June 2022, largest increase in 40 years. Available at: <https://www.bls.gov/opub/ted/2022/consumer-prices-up-9-1-percent-over-the-year-ended-june-2022-largest-increase-in-40-years.htm> (Accessed: 13 Aug 2022)

CNBC (2022). Are tech giants ‘sunsetting’? Strategists warn of big tech under pressure. Available at: <https://www.cnbc.com/2022/02/14/us-china-tech-giants-under-regulatory-pressure-competition.html> (Accessed: 27 Jul 2022)

Huijian Dong, Xiaomin Guo, Han Reichgelt & Ruizhi Hu (2020). Predictive power of ARIMA models in forecasting equity returns: a sliding window method. J Asset Manag 21, 549–566 (2020). <https://doi.org/10.1057/s41260-020-00184-z>

IBM (2020). Neural Networks. Available at: <https://www.ibm.com/cloud/learn/neural-networks#toc-how-do-neu-vMq6OP-P> (Accessed: 14 Aug 2022)

Xuan Ji, Jiachen Wang, Zhijun Yan (2021). A stock price prediction method based on deep learning technology. ISSN: 2398-7294.

Calomiris C.W., Mamaysky H. (2018). How news and its context drive risk and returns around the world. Journal of Financial Economics. 2018;133:299–336.