Emerging from the Pandemic: A Study on Global Stock Index Return Trend Using a Hybrid Approach of Linear Regression and LSTM Models

Jinghan Ji[*], Jingxi Feng[†], Shengran Huang[‡], Xutong Zhu[§]

Apr, 2024

Abstract

The sudden outbreak of the COVID-19 pandemic has brought about an undeniable impact on the national economy. In order to delve into how to more efficiently prevent the risks associated with major public health emergencies in the future, this paper focuses on the stock market's return rate. By integrating a multiple linear regression model with an LTSM model, a comparative assessment of the stock market's impact before and after the pandemic is conducted. The research findings indicate that there is no significant difference in stock market return rates between the two periods. Furthermore, a comparison is made between the improved model and the Purely Linear Regression Model, ARIMA, Prophet, and ESM, revealing that the improved model provides more accurate predictions of stock market return rates. This discovery holds the potential to significantly reduce the latent risks brought about by major events and crises, enabling governments to make relevant policy adjustments based on the independent variables mentioned in this paper, thus mitigating potential economic losses.

Key words: Stock Index Returns, COVID-19 pandemic, Post-Pandemic Dynamics, Trend Prediction, Long Short-Term Memory (LSTM), Regression Analysis.

[*]Business School, University of Edinburgh, Edinburgh, EH3 9DH, United Kingdom, jijinghan05@gmail.com

[†]"University of California, Los Angeles" , Address: 11050 Strathmore Drive, 90024. kinkikinsey@gmail.com

[‡]"Beijing Forestry University" , Address: Qinghua East Road No.35-5, Haidian Beijing China. huangshengran530@gmail.com

[§]"University of California, San Diego" , Address; 9500 Gilman Dr, La Jolla, CA 92093, USA. zxtcjs@gmail.com

1 Introduction

The COVID-19 pandemic, which swept across the globe with unprecedented speed and scale, has not only upended the fabric of daily life but has also left indelible marks on the global economic landscape. The measures implemented to curb the virus's transmission, including nationwide quarantines and lockdowns, have given rise to a myriad of consequences, which ripple through various sectors of the economy (Uchehara et al., 2020). Business closures reduced commercial activities, and disruptions to global supply chains have ushered in a new era of uncertainty, where economic equilibrium teeters precariously. While these pandemic-induced changes have left an indelible mark on nearly every aspect of human existence, the financial sector has borne witness to some of the most profound transformations.

The stock market, as a barometer of economic health and vitality, has experienced fluctuations of unparalleled magnitude and unpredictability. The COVID-19 pandemic has compelled a thorough re-evaluation of the traditional determinants that have historically shaped stock indices. In particular, the macroeconomic parameters that once wielded significant influence over market dynamics have found themselves in a shifting landscape, where new patterns emerge and conventional wisdom is called into question.

In this dynamic context, our research embarks on a comprehensive exploration of the profound effects that the COVID-19 pandemic has imposed on stock indices. We delve into a multi-faceted analysis, considering both pre- and post-pandemic periods, seeking to unearth the evolving relationships between various macroeconomic determinants and stock index returns. Our investigation spans five high-GDP countries—namely, the United Kingdom, the United States, China, Germany, and Japan—chosen for their universal relevance and global representativeness in the face of unprecedented crises.

While the pandemic's impact on the global stock market is palpable, there remains a notable dearth of comprehensive analyses that scrutinize this impact from a broader, macroeconomic perspective. Our research aims to bridge this knowledge gap by examining the intricate dynamics within the financial domain. As we navigate this new and uncharted territory, the findings of this study promise to contribute to a deeper understanding of the stock market's resilience and responsiveness in the face of external disruptions.

The methodology of our research comprises two components, the first of which focuses on the influence of six key parameters on stock index return rates. The dataset we collected is crafted from key economic indicators, and a linear regression model is tailored to assess the pre- and post-pandemic relationships. The second component introduces Time Series Long Short-Term Memory (LSTM) techniques to enhance trend prediction, thereby minimizing prediction errors and improve the model's resilience during unforeseen disruptive events. Additionally, a comparative analysis with existing models was conducted to ascertain the accuracy and effectiveness of our prediction model.

The COVID-19 pandemic has reshaped our world, and through this research, we seek to analyze the interplay between global crises and financial markets, thereby provide valuable information on potential strategies for investors, analysts, and policymakers in the post-pandemic era.

2 Literature Review

There are some literatures surrounding stock return predictability to give insights about the factors influencing stock market behavior. For example, several studies emphasize the significance of economic indicators and time-varying risk premia, which align with the objectives of this research. The incorporation of micro and macro-economic variables, as suggested by Thenmozhi (2006), also resonates with our approach.

Marquering and Verbeek (2004) lay a foundational framework for understanding the dynamics of stock return predictability. Their work emphasised the positive relationship between high economic volatility and return predictability. This finding reveals role of incomplete learning following significant economic shocks and time-varying risk premia. This insight rationalizes the incorporation of economic indicators and temporal segmentation in our methodology, enabling a more nuanced analysis of return trends.

Boudoukh and Richardson (1993) provide compelling evidence for a positive correlation between nominal stock returns and inflation at extended horizons. Their findings emphasize the intricate relationship between macroeconomic factors and stock market performance.

Gonzalo and Taamouti (2017) undertake a thorough examination of how both anticipated and unanticipated unemployment rates affect the distribution and quantiles of stock prices. Their results challenge conventional wisdom, showing that only the anticipated unemployment rate has a significant impact on stock prices. This outcome is explained by the Fisher and Phillips curve equations, which suggest that a high unemployment rate prompts intervention from the Federal Reserve, subsequently resulting in higher stock market prices. This underscores the significance of comprehending the predictive nature of economic indicators.

Alexius and Spång (2018) discover a substantial correlation between stock prices and the domestic GDP. Their estimates indicate that a one percent increase in stock prices is linked to a greater than one percent rise in GDP. This observation highlights a positive relationship between the performance of the stock market and economic growth.

Liow (2006) offers three significant observations: initially, there is an immediate and enduring connection between the stock market and prices of residential and office properties; secondly, the influence of combined residential and office property prices on stock market prices diminishes when adjustments are made for shifts in macroeconomic conditions; and lastly, while the long-term impact of the stock market is predominantly steered by the office property market, the residential property market wields a more substantial influence in the short run. These findings impart crucial nuances for comprehending the interaction between property markets and stock market performance.

Some literatures also highlight the potential of alternative modeling techniques. For instance, Boyacioglu and Avci (2010) introduce the potential of the ANFIS model in financial applications. Although not directly aligned with our methodology, their work serves as a reminder of the diverse range of techniques available for financial modeling, encouraging consideration of alternative methodologies and justifying the incorporation of LSTM techniques in our study.

Qiu et al. (2016) introduces a hybrid methodology that combines global search techniques with neural networks, demonstrating its effectiveness in overcoming limitations associated with gradient search algorithms. While not directly aligned with our approach, this study emphasizes the value of innovative techniques in refining prediction models. This further validates our decision to integrate regression and LSTM methodologies to enhance prediction accuracy.

Chen et al. (2015) utilize LSTM neural networks for stock return prediction, emphasizing the impact of normalization and market indexes on prediction accuracy. Their findings underscore the importance of data preprocessing techniques, which align closely with our rigorous data processing protocol. Additionally, their exploration of different stock sets highlights the necessity for a nuanced analysis, affirming our strategy of considering high-GDP countries.

Hansson (2017) employs LSTM neural networks to forecast financial returns, with a specific focus on the Swedish, US, and Brazilian stock markets. The study suggests variations in predictability across these markets, indicating potential differences in data-driven influences. This observation further strengthens our rationale for concentrating on high-GDP countries with broad global relevance and representation.

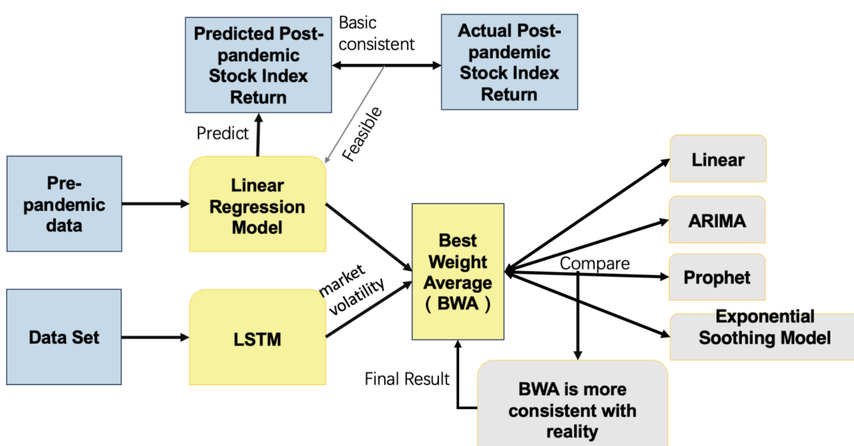
As highlighted by Kalpakis, Gada, and Puttagunta (2001), traditional regression approaches frequently fall short in fully capturing the intricate dynamics inherent in time series data. Autoregressive Integrated Moving Average (ARIMA) models, on the other hand, stand out as a robust alternative, demonstrating effectiveness in aptly modeling a diverse array of environmental and socioeconomic time-series datasets. Notably, Garlapati et al. (2021) leveraged the Facebook Prophet model for stock price prediction. The other model: exponential smoothing method is relatively simple but robust approaches to forecasting (Billah et al., 2006). In the present research, we opt to employ these established models, including ARIMA, Facebook Prophet, and exponential smoothing model to rigorously test and compare the effectiveness of our proposed model against these.

In summary, while existing literature provides valuable insights into stock return predictability, there remains a gap in comprehensive analyses, especially in the aftermath of significant economic shocks like the Covid-19 pandemic. This study's methodology, which combines regression analysis and LSTM techniques, aims to fill this gap by offering a nuanced approach to understanding stock market dynamics in both pre- and post-pandemic periods. The incorporation of key economic indicators, temporal segmentation, and innovative modeling techniques positions this research to make valuable contributions to the field of financial analysis.

3 Methodology

The methodology employed in this study aims to develop a robust predictive model for Global Stock Index Return Trend that accommodates both pre and post-pandemic scenarios. Initially, a comprehensive set of parameters, including annual GDP growth rate, interest rate, inflation rate, unemployment rate, residential property prices, and loans to the private sector, was utilized to construct a linear regression model. To enhance the precision of the predictive model, the analysis was extended to incorporate Long Short-Term Memory (LSTM) models.

Subsequently, a Best Weight Average mechanism was introduced to synergize the strengths of the linear regression and LSTM models. The resulting integrated model was then compared with the conventional ARIMA model, Facebook Prophet and exponential smoothing model, demonstrating accuracy in forecasting Stock Index Returns. This methodology section outlines the sequential steps undertaken to formulate a predictive framework capable of navigating the complexities introduced by the pandemic and providing more accurate forecasts compared to existing models. The specific process is shown in flow chart 1.



Flowchart 1

Part 1 Influence of Six Key Parameters on the Stock Index Return Rate

1.1 Sample Selection and Data Processing

In this study, we curated a dataset comprising key economic indicators from five high-GDP nations spanning the years 2000 to 2023. These parameters encompassed the annual GDP Growth Rate (expressed in percentage), Interest Rate (also in percentage), Inflation Rate (averaged quarterly), Unemployment Rate (in percentage), Residential Property Prices and Loans to Private Sector. The deliberate choice of high-GDP countries (UK, US, China, Germany, and Japan.) was motivated by the desire to examine mature markets with universal relevance and global representativeness when assessing the repercussions of crises on global stock indices. As the dependent variable, we collected data on the 'Index Return Rate', specifically focusing on the Nikkei 225, FTSE All-Share Index, DAX Index, Shanghai Composite index, and S&P index.

The temporal domain was divided into two distinct segments: data from 2000 to 2019 as pre-COVID, while data from 2020 to 2023 as the post-COVID.

The rigorous data processing protocol included the following steps:

- Removal of empty rows, ensuring a pristine dataset.
- Conversion of rate variables from percentage to decimal for uniformity and precision in subsequent analyses.
- In our analysis of 'loans to the sector' across five distinct countries, we encountered the challenge of disparate currencies and potential variations in loan magnitudes. To address this and enable meaningful cross-country comparisons, we employed a data standardization method. Specifically, for each country, we calculated the mean and standard deviation of its loan amounts. Subsequently, every individual loan data point within a country was standardized by subtracting its country-specific mean and then dividing by its country-specific standard deviation. This transformed each loan value into a measure of how many standard deviations it lies from the country's mean, effectively nullifying the original currency units and magnitudes.

1.2 Model Training

The dataset was systematically divided into pre- and post-pandemic periods, a crucial step in developing a robust linear regression model aligned with our research objectives.

To uphold the model's precision and applicability, a meticulous partitioning strategy was implemented.

Specifically, data from the pre-COVID era was stratified, with 70% allocated for training purposes and the remaining 30% reserved for validation, ensuring a rigorous assessment of the model's performance. This division played a pivotal role in enabling a thorough iterative analysis using Python. The primary objective was to ascertain the polynomial degree that minimizes the Mean Squared Error (MSE) within a maximum of 10 iterations. Once identified, this optimal degree was utilized to construct the regression model formally. This model underwent training using the pre-COVID data and was subsequently applied to forecast outcomes in the post-COVID dataset. This process shed light on the pandemic's impacts on the stock indices domain. The selected model demonstrated a polynomial degree of 1, indicating a linear regression model well-positioned to provide robust predictive capabilities.

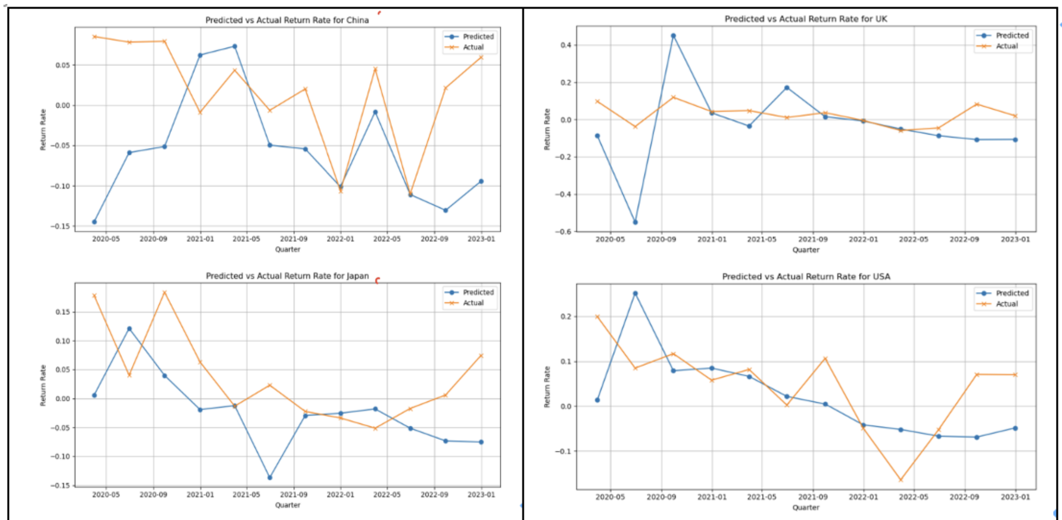
1.3 Evaluate the Model Performance

We performed polynomial regression on the data, specifically focusing on the 'After COVID' period, and then evaluating the model's performance using the MSE. The MSE after covid was found to be around 0.01133, which means model's proficiency and ability to explain the data.

1.4 Differences Between Actual and Predicted

To facilitate time series analysis and forecasting, we initially undertook a series of operations to reformat the temporal data. This involved converting the "time" column into a datetime format and subsequently designating it as the index for the DataFrame. Notably, for quarters denoted as 'Q1,' 'Q2,' 'Q3,' and 'Q4,' the corresponding dates were set as March 31st, June 30th, September 30th, and December 31st, respectively, of the respective year. This standardization of date format enabled a uniform approach for time series analytics and predictions.

Subsequently, a comparative analysis was conducted by comparing the predicted and actual return rates for each country, as illustrated in Figure 2. Evidently, except for the United Kingdom, a conspicuous opposite trend materializes in the comparison of actual and predicted values across multiple nations in the initial half of 2020. This observation prompts a contemplation of whether this phenomenon can be attributed to the disruptive impact of the Covid-19 pandemic.



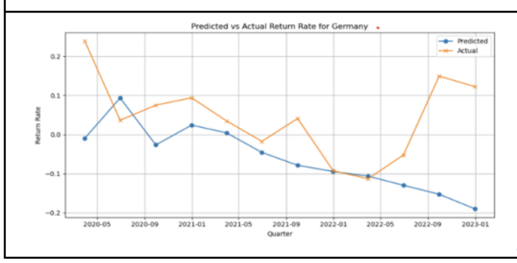


Figure 2. comparison between prediction and actual return rate each country

Then we Draw the Difference between Prediction and Actual values for each country shown on figure 3. The difference in return rate on y-axis indicates “Prediction minus actual stock indices return rate”. We can intuitively find that especially in early 2020 and Q2 of 2022, the predictionactual deviation values among five countries are relatively large, which shows a mighty degree of volatility in data. This phenomenon raises the potential need for improved models.

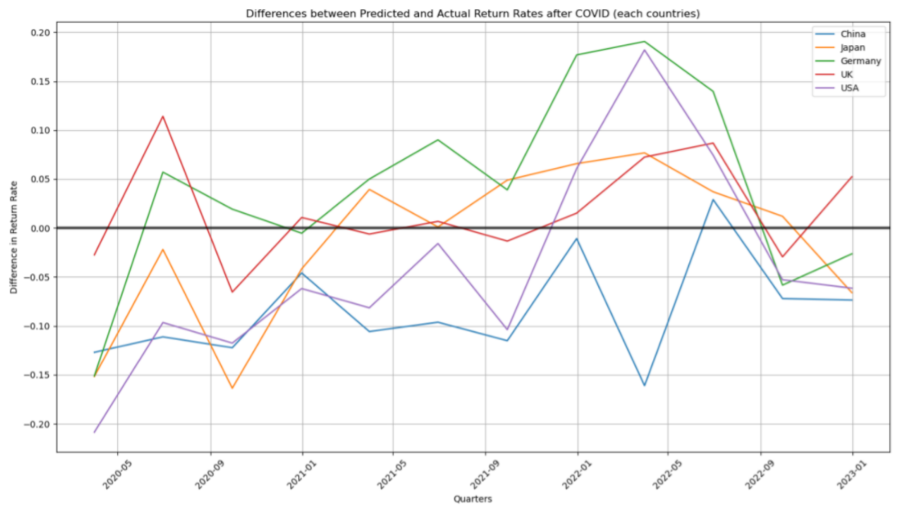


Figure3 The Difference between Prediction and Actual Values for each country

Following the derivation of the average prediction globally based on regression results from five countries, we proceed to compare these average predictions with actual values, as illustrated in Figure 4. By amalgamating the prediction-actual deviation values across the five countries, noteworthy trends emerge. Specifically, during the first half of 2020, in the intervals spanning July 2021 to October 2021, and post-January 2022, we observe an inverse trend or a notable difference between the average forecasted values and the average actual values. Once again, this phenomenon underscores the potential necessity for enhanced models.

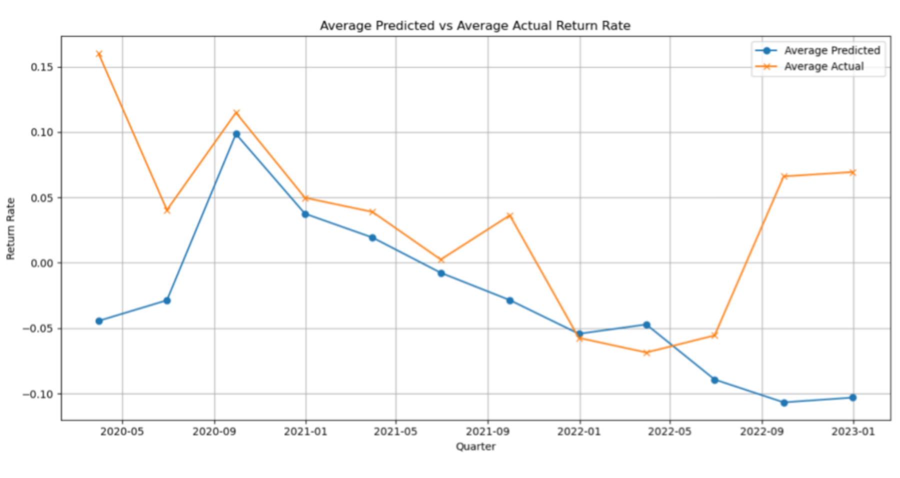


Figure4 The Difference between Average Prediction and Average Actual Values

We conducted a global analysis, employing a T-test and examining P-values for mean differences. The results indicated a t-statistic of -1.3226415359646189 and a p-value of 0.19105864043505913 (which exceeds the 0.05 significance level). This suggests that there is no statistically significant evidence to indicate that the parameters influencing the return rate differ before and after COVID-19, leading to a failure to reject the null hypothesis. However, it is important to note that the failure to reject the null hypothesis does not serve as proof for the alternative hypothesis. Instead, it signifies that, based on the available data, the two periods (pre and post COVID-19) do not exhibit statistically significant differences in the context of return rates.

1.5 Anova & P-values & R-Squared for model validation

We examine the Anova & P-values & R-Squared for model validation, and the regression result as followed shown on figure 5.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:            Return rate   R-squared:                       0.054
Model:                            OLS   Adj. R-squared:                  0.038
Method:                 Least Squares   F-statistic:                     3.492
Date:                Mon, 30 Oct 2023   Prob (F-statistic):            0.00227
Time:                        15:26:24   Log-Likelihood:                 305.21
No. Observations:                 375   AIC:                            -596.4
Df Residuals:                     368   BIC:                            -568.9
Df Model:                           6
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         -0.0050      0.020     -0.245      0.807      -0.045       0.035
x1             0.4720      0.194      2.436      0.015       0.091       0.853
x2            -0.5249      0.368     -1.428      0.154      -1.248       0.198
x3            -1.3720      0.464     -2.958      0.003      -2.284      -0.460
x4             0.5713      0.316      1.806      0.072      -0.051       1.193
x5             0.0008      0.001      0.742      0.459      -0.001       0.003
x6            -0.0064      0.008     -0.763      0.446      -0.023       0.010
==============================================================================
Omnibus:                       21.526   Durbin-Watson:                   1.886
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               65.511
Skew:                           0.052   Prob(JB):                     5.95e-15
Kurtosis:                       5.045   Cond. No.                         615.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

Figure 5 Regression Result

For Model Fit, R-squared (R^2) is 0.054, indicates that the model explains approximately 5.4% of the variance in the dependent variable, so there is no overfitting. For Statistical Significance, the F-statistic is 3.492, Prob (F-statistic) is 0.00227. The F-statistic tests the overall significance of the model. The low p-value (0.00227) suggests that at least one of the independent variables significantly impacts the dependent variable.

Under the coefficients column, ×1 to ×6 indicates GDP growth rate, Interest rate, Inflation rate, Unemployment rate, Residential property prices, Loans to private sector respectively. For Significance of Coefficients: The 'P>|t|' column represents the p-value associated with each coefficient. A low p-value indicates that the coefficient is statistically significant. Confidence Intervals: The [0.025 0.975] columns provide the 95% confidence interval for each coefficient. This range suggests the plausible values for the true population parameter.

Overall, this regression model attempts to explain the variation in the dependent variable using the specified independent variables. It is important to note that while the model shows some explanatory power, its overall fit may still be limited, as indicated by the relatively low R-squared value. Further analysis and refinement of the model may be warranted.

1.6 View the Equation

The regression equation representing the return rate is as follows:
Return rate = -0.0050 + 0.4720(GDP growth rate) + -0.5249(Interest rate) + -1.3720(Inflation rate) + 0.5713(Unemployment rate) + 0.0008(Residential property prices) + -0.0064(Loans to private sector)

1.7 Effect Size of Parameters

The effect sizes derived from the model shed light on the magnitudes of the relationships between predictors and the return rate. Figure 6 reveals that the GDP growth rate (with an effect size of 1.59), unemployment rate (1.14) and residential property prices (4.59) demonstrates a moderately positive correlation with the return rate. Conversely, the Interest Rate (-1), Inflation rate (-2.02), Loan to Private Sector (-4.87) demonstrates a negative effect.

```
const                           9.437784e-03
const                          -1.276503e-17
GDP growth rate                 1.585672e-02
Interest rate                  -1.000759e-02
Inflation rate                 -2.018432e-02
Unemployment rate               1.142570e-02
Residential property prices     4.586875e-03
Loans to private sector        -4.872815e-03
dtype: float64
```

Figure6 Effect Size of Parameters

Part 2 Prediction combine previous model with time series (to minimize prediction error in return rate trend when encounter pandemic like Covid 19)

2.1 Selection of LSTM

Initially, we trained a model based on the S&P500 index data spanning from 2001 to 2022. The rationale behind selecting the S&P500 index lies in its inclusion of globally recognized Fortune 500 companies. Subsequently, we transformed the stock index data, assigning a value of 1 to an increase in the stock index and 0 to a decrease. This transformation enabled us to visually compare actual and predicted trends in monthly return rates, as depicted in Figure 7. On Figure 7, the red dots represent predicted increases, and the blue dots represent actual monthly return. Red dots on the bottom of the blue line means the successful of increase trend prediction.

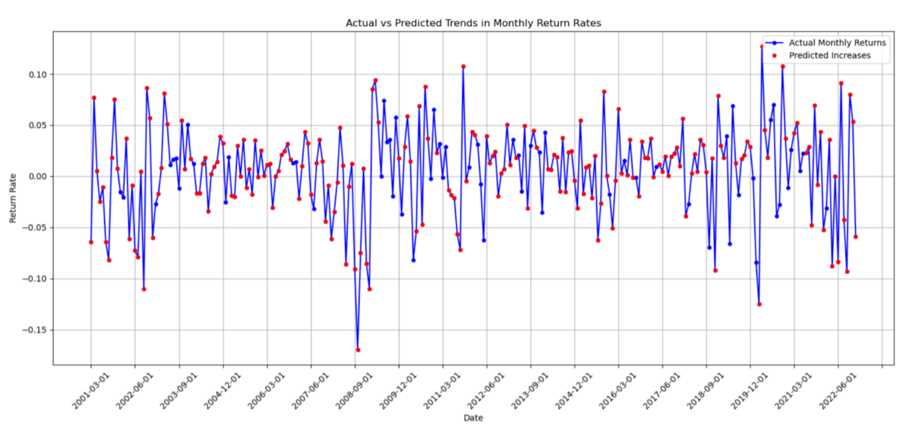


Figure7 Actual vs Predicted Trends in Monthly Return Rates

2.2 Purely looking at prediction with LSTM (with no regression model)

First, we draw out the average prediction of increase & decrease in return rate from only time series shown on figure 8. Figure 8 shows market volatility and possible price ranges or rises and falls. It should be noted that LSTM prediction cannot be used directly to make predictions. It only shows the upper and lower fluctuations in return rates. To get a more appropriate model, we need to combine the LSTM model with the regression model previously obtained in part 1.

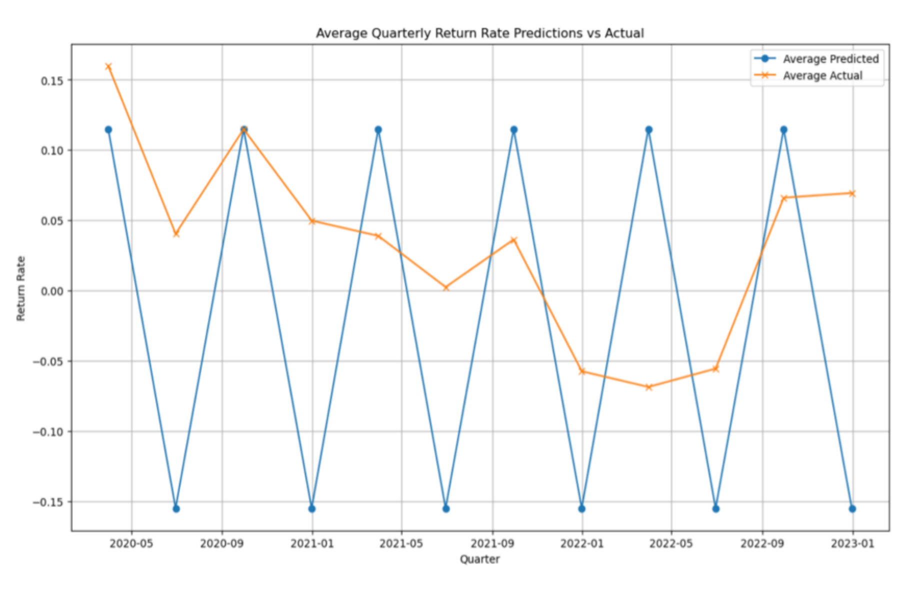


Figure8 Average Quarterly Return Rate Predictions vs Actual

2.3 Enhance Trend Prediction Combine Regress Model & LSTM Machine Learning

We calculated the best weight average with less MSE to get the best weight average. The optimized combination equation is: Combined Prediction = 0.7511 * Linear Regression Prediction + 0.2489 * LSTM Prediction

Then we plot out the comparison between regression model prediction (green line), actual (black line) and weighted average prediction (red line) (0.7511 * Linear Regression Prediction + 0.2489 * LSTM Prediction) illustrated on figure 9. It can be seen that the weighted forecast represented by the red line is more realistic, the deviations are only appeared in the trends of Q1, Q2 and Q4 of 2022.

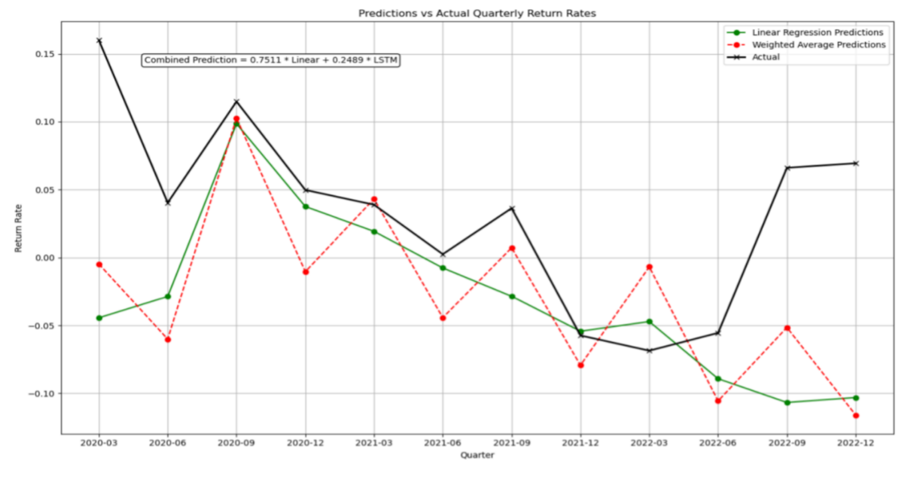


Figure9 comparison between regression model prediction,actual and weighted average prediction

In order to further compare whether the new weighted prediction is more accurate than the original Linear Regression model, we draw a bar chart (figure 10) for seeing performance on each method. This is 10% more accuracy of trend by using this combined trend prediction.

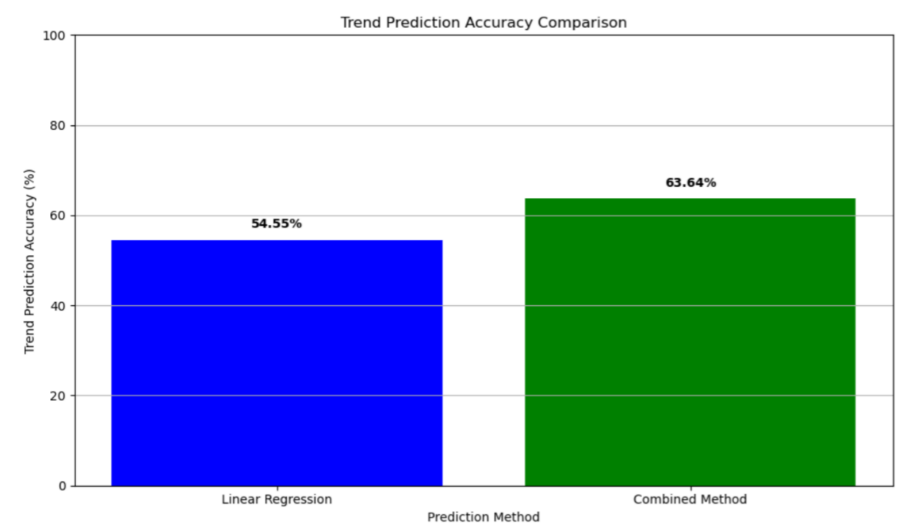
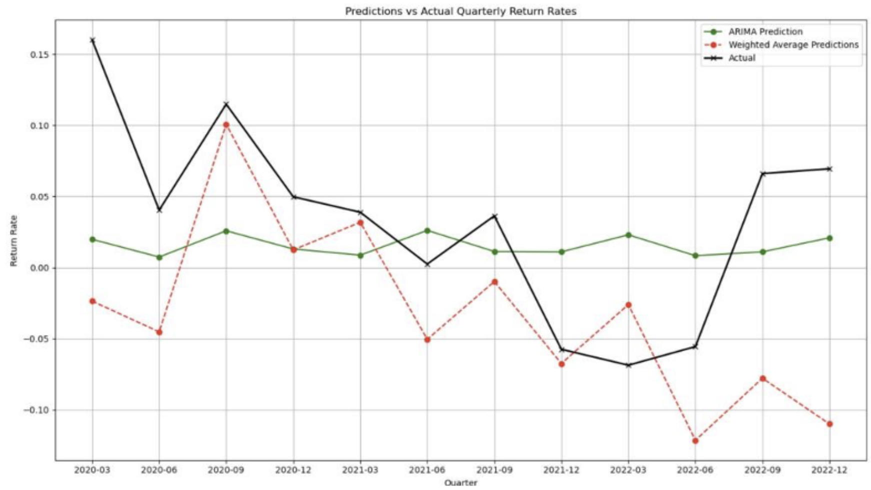
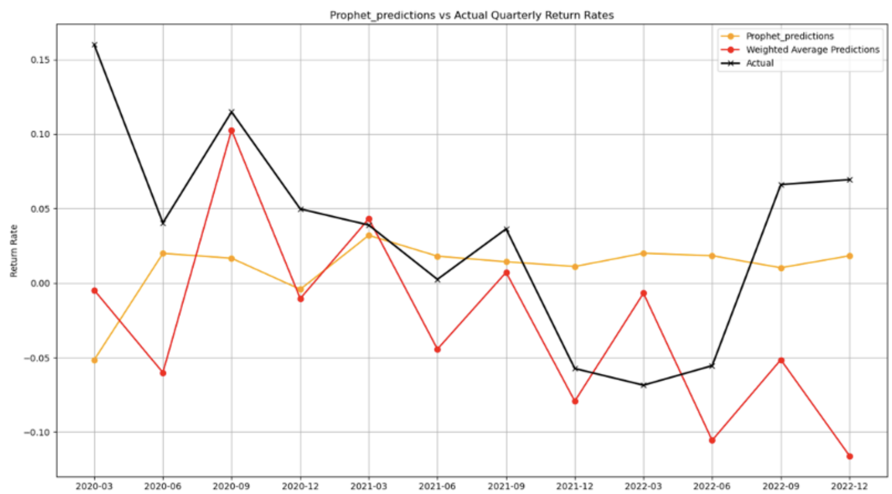


Figure10 Trend Prediction Accuracy Comparison

To validate the efficacy of our predictions, we conducted a comparison with established models, namely ARIMA, Facebook Prophet, and the exponential smoothing model (illustrated in Figure 11). Our predictions closely align with the actual stock index return rates.





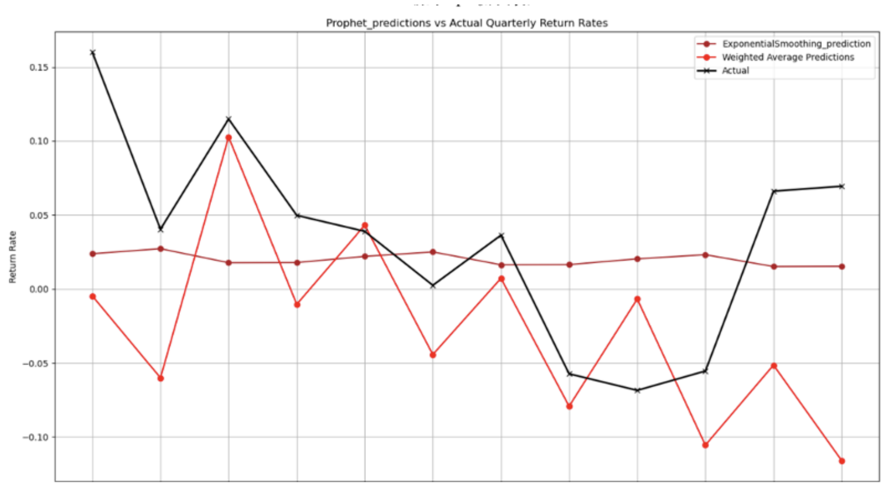


Figure11 Comparison to 3 models

4 Results

This research presents a comprehensive analysis of the intricate relationship between key economic parameters and stock index returns, with a specific focus on the pre- and post-Covid-19 landscape. The study comprises two primary components: the influence of six pivotal parameters on stock index return rates, analyzed through a trend prediction regression model, and the integration of Time Series Long Short-Term Memory (LSTM) techniques to enhance prediction accuracy in return rate trends. The dataset spanning five high-GDP nations from 2000 to 2023 underpins the research, with a focus on mature markets. Our investigation centers on the Nikkei 225, FTSE All-Share Index, DAX Index, Shanghai Composite Index, and S&P Index as dependent variables.

Part 1 Influence of Six Key Parameters on the Stock Index Return Rate

The dataset, divided into pre- and post-Covid-19 periods, underwent thorough data processing and model training. The model achieved a polynomial degree of 1, indicating a linear regression model suited for robust predictive capabilities. The Mean Squared Error (MSE) after Covid-19 was approximately 0.01133, signifying the model's proficiency. However, a significant deviation between actual and predicted values in early 2020 raised questions about the pandemic's impact, which is worth to analyze in the future research. The T-test and p-values of means differences globally indicated that there's no statistically significant evidence suggesting differences in return rates before and after Covid-19.

The Anova & P-values & R-Squared for model validation revealed an R-squared of 0.054, indicating that the model explains approximately 5.4% of the variance in the dependent variable. The F-statistic and p-value suggested that at least one of the independent variables significantly impacts the dependent variable. The effect sizes of parameters demonstrated various magnitudes of relationships between predictors and the return rate, offering valuable insights into the relative influence of each variable, which is also a good way for making policies by government.

The regression equation representing the return rate is as follows:

$$\text{Return Rate} = \text{-}0.0050 + 0.4720 \cdot x_1 - 0.5249 \cdot x_2 - 1.3720 \cdot x_3 + 0.5713 \cdot x_4 + 0.0008 \cdot x_5 - 0.0064 \cdot x_6$$

$where$
$x_1 = GDPGrowthRate$
$x_2 = InterestRate$
$x_3 = InflationRate$
$x_4 = UnemploymentRate$
$x_5 = ResidentialPropertyPrices$
$x_6 = LoantoPrivateSector$

Part 2 Prediction Combining Regression Model and LSTM Techniques

Time series analysis using LSTM was employed initially, showing its proficiency in simulating fluctuations in return rates. However, this prediction alone had limitations, prompting the combination of LSTM and the regression model to enhance trend prediction. An optimized combination equation was derived to achieve the best weight average for predictions. The comparison between the regression model, actual values, and the weighted average prediction demonstrated that the weighted forecast was more realistic, with deviations primarily occurring in specific quarters of 2022.

The performance comparison between the combined method and the original Linear Regression model showed approximately 10% greater accuracy using the combined approach.

The optimise combination of stock indix return rate= 0.7511 * Linear Regression Prediction + 0.2489 * LSTM Prediction

5 Discussion

The research findings offer substantial insights into the complex relationship between economic parameters and stock index returns, with a particular emphasis on the Covid-19 pandemic's impact. While the linear regression model explained a limited portion of the variance in the dependent variable, it served as a foundational element for subsequent analysis. The combination of LSTM techniques with the regression model showcased the potential to improve prediction accuracy and enhance trend forecasting. This integrated approach outperformed the regression model alone, particularly during quarters with notable deviations.

However, it is important to acknowledge the limitations of the research. The Covid-19 pandemic is a multifaceted event influenced by numerous factors beyond economic parameters. As such, while this study offers valuable insights, it cannot capture the entirety of the pandemic's impact on stock indices. Additionally, the financial domain is highly dynamic and influenced by various exogenous factors that extend beyond the scope of this research.

6 Conclusion

In conclusion, this study focused on developing a robust predictive model for Stock Index Returns that is applicable to both pre and post-pandemic periods. While the comparison (figure 11) highlights the effectiveness of the proposed model, it is crucial to acknowledge that the accuracy and efficacy of the model may be subject to further refinement through continued research. Future studies should explore additional factors and consider evolving market conditions to ensure the continued relevance and reliability of the predictive framework in dynamic financial landscapes. This research lays a foundation for ongoing investigations aimed at refining and advancing predictive models for Stock Index Returns in the face of changing economic environments.

The findings of this study are valuable for investors, analysts, and policymakers seeking a deeper understanding of stock market dynamics in the wake of external disruptions, such as the Covid-19 pandemic. The combination of regression analysis and LSTM techniques demonstrated the potential to improve prediction accuracy and provide more realistic trend forecasts.

Future research can focus on refining the LSTM models used in this study. Incorporating more sophisticated neural network architectures and fine-tuning hyperparameters can potentially enhance prediction accuracy. Additionally, exploring the use of alternative deep learning models, such as recurrent neural networks (RNNs) and transformers, may provide new insights of enhancing the prediction accuracy.

Incorporating sentiment analysis of news, social media, and other textual data can provide a deeper understanding of market sentiment and investor behavior. Natural language processing (NLP) techniques can help gauge market sentiment and its impact on stock returns (Yıldırım et al., 2018).

While this study focused on the Covid-19 pandemic, future research can explore the impact of other global economic events on stock market behavior. Events like financial crises, geopolitical tensions, and major policy changes can have profound effects on financial markets. Analyzing how different events influence stock indices and refining predictive models accordingly is a promising direction.

Reference

Alexius, A., & Spång, D. (2018). Stock prices and GDP in the long run. Journal of Applied Finance & Banking, 8(4), 1792–6599.

Billah, B., King, M. L., Snyder, R. D., & Koehler, A. B. (2006). Exponential smoothing model selection for forecasting. International journal of forecasting, 22(2), 239-247.

Boyacioglu, M. A., & Avci, D. (2010). An Adaptive Network-Based Fuzzy Inference System (ANFIS) for the prediction of stock market return: The case of the Istanbul Stock Exchange. Expert Systems with Applications, 37(12), 7908–7912. https://doi.org/10.1016/j.eswa.2010.04.045

Garlapati, A., Krishna, D. R., Garlapati, K., Rahul, U., & Narayanan, G. (2021, April). Stock price prediction using Facebook Prophet and Arima models. In 2021 6th International Conference for Convergence in Technology (I2CT) (pp. 1-7). IEEE. Gonzalo, J., & Taamouti, A. (2017). The reaction of stock market returns to unemployment. Studies in Nonlinear Dynamics & Econometrics, 21(4), 20150078.

Hansson, M. (2017). On stock return prediction with LSTM networks.

K. Chen, Y. Zhou and F. Dai, "A LSTM-based method for stock returns prediction: A case study of China stock market," 2015 IEEE International Conference on Big Data (Big Data), Santa Clara, CA, USA, 2015, pp. 2823-2824, doi: 10.1109/BigData.2015.7364089.

Kalpakis, K., Gada, D., & Puttagunta, V. (2001, November). Distance measures for effective clustering of ARIMA time-series. In Proceedings 2001 IEEE international conference on data mining (pp. 273-280). IEEE.

Liow, K. H. (2006). Dynamic relationship between stock and property markets. Applied Financial Economics, 16(5), 371–376.

Marquering, W., & Verbeek, M. (2004). The Economic Value of Predicting Stock Index Returns and Volatility. Journal of Financial and Quantitative Analysis, 39(2), 407–429. https://doi.org/10.1017/s0022109000003136

Qiu, M., Song, Y., & Akagi, F. (2016). Application of artificial neural network for the prediction of stock market returns: The case of the Japanese stock market. Chaos, Solitons & Fractals, 85, 1–7. https://doi.org/10.1016/j.chaos.2016.01.004

Thenmozhi, M. (2006). Forecasting stock index returns using neural networks. Delhi Business Review, 7(2), 59-69.

Uchehara, I., Hamma-adama, M., Obiri, K.A., Jafarifar, N. and Moore, D. (2020). Impacts and Risk Management of COVID-19 Pandemic on Real Estate Supply Chain. International Journal of Real Estate Studies, [online] 14(S1), pp.41–53. Available at: https://intrest.utm.my/index.php/intrest/article/view/62/28.

Yıldırım, S., Jothimani, D., Kavaklıoğlu, C., & Başar, A. (2018, December). Classification of" hot news" for financial forecast using NLP techniques. In 2018 IEEE International Conference on Big Data (Big Data) (pp. 4719-4722). IEEE.