ISSN: 2997-2256

Research Article

Journal of Investment, Banking and Finance

Leveraging Machine Learning Techniques to Study The Stock Market Dynamics in India

Prokarsha Kumar Ghosh*

Actuarial Science, Institute of Actuaries of India, India

***Corresponding Author**
Prokarsha Kumar Ghosh, Actuarial Science, Institute of Actuaries of India, India.

Submitted: 2025, Feb 07; **Accepted:** 2025, Feb 28; **Published:** 2025, Mar 10

Citation: Ghosh, P. K. (2025). Leveraging Machine Learning Techniques to Study The Stock Market Dynamics in India. *J Invest Bank Finance, 3*(1), 01-07.

Abstract
This abstract explores the application of time series analysis, machine learning, and deep learning methods in forecasting long-term stock market trends in the Indian financial market. Even though short-term forecasting is challenging due to the complexities of the markets and the emotional factors it has involved, long-term trends become more predictable in data science techniques. Recent studies have been able to showcase that time series analysis happens to be an effective technique in recognizing patterns and predicting future movements based on historical data. One may also observe the inter-stock relationships and understand how those relate to market trends or potential investments. Ongoing advancements in statistical and analytical methods are driving significant progress in stock market analysis. As machine learning techniques keep improving, they will help make forecasting models better, giving investors and financial experts more useful information to make smarter decisions. These developments are poised to improve stock market prediction and offer crucial insights into market behaviour, benefiting stakeholders in the financial sector.

Keywords: Indian Stock Market, Statistical Analysis, Time Series Analysis, ANOVA, ARIMA, ARCH, GARCH, LSTM Model.

1. Introduction
Investing in stock markets is one of the greatest inventions and a major input on the global economy for years. Investing in the stock market and achieving high profits with low risks, investors are using many statistical techniques and many analytical approaches to make decisions in financial markets and it takes an increasingly pre-eminent place in the global economy for years. Predicting the stock market trend is a big challenge. Analysing the trend of the stock market price is one of the trickiest calculations. There are many theories regarding the stock markets that have been published for years. According to Yang et al., 2019, the analysis and prediction of the trend are important measurement tools to evaluate the risk in the stock market data, which is significant to investors in the market and government regulators [1]. Therefore, to analyse the risk in price valuation in the market and identify the level of resistance for buyers and sellers' volatility modelling has been used. Volatility modelling is also used to detect price fluctuation in the financial market. In modern days there are many technological tools and new analytical methods are used to predict future stock prices.

This paper has discussed on volatility and complexity of Indian financial markets. Therefore, market risk prediction is highly complex for the global economy and not easy to avoid. Many experts and professionals in this market, consider S&P BSE Energy to be an inadequate representation of many wealthy companies in the Indian stock market which are compared to one of the largest broader in Indian stock markets such as the S&P Bankex. In the context of the Indian financial market S&P BSE Energy is designed to provide its investors with a benchmark reflecting companies included in the S&P BSE BANKEX that are classified as members. This paper reviews the association between two stock indices S&P BSE Energy and S&P BSE BANKEX. This means one unit price change in S&P BSE Energy how much affects S&P BSE BANKEX. A comparative analysis between the S&P BSE Energy Index and the S&P BSE All Cap Index was conducted using linear regression, ANOVA, ARIMA ARCH, GARCH and LSTM model. A strong positive correlation between the indices was identified through linear regression, indicating similar movements.

ANOVA results showed no significant difference in returns, suggesting comparable volatility and returns. ARIMA models forecasted similar upward trends with typical market fluctuations for both indices. In conclusion, the analysis revealed similar patterns in returns and volatility, highlighting the energy sector's significant influence on overall market trends. ARCH model is used to analyse time-varying volatility in stock returns, while GARCH model used to extend analysed ARCH model by capturing volatility persistence. On the other hand, LSTM is used to identify complex, non-linear relationships between stocks, making them useful for forecasting and understanding dependencies in stock behaviour. Together, these models provide a comprehensive view of stock dynamics. Therefore, in this case study, statistical analysed is used to observe the association between two stock indices in the context of the Indian financial market.

2. Literature Review

Analysing the market trend and modelling the market volatility is one of the most important measurements to examine the trend of the market and calculate the risk in investing in the market. Therefore, forecasting the volatility of market data is more significant to its investors and the government regulatory authority. Many authors and researchers suggested that stock markets are a random walk and cannot be predicted. Many authors also suggested that stock price predicting is too difficult to find the behavioural pattern of the market. Analysing the trend is unpredictable in short term but there is a chance to predict the future outcome in the longer term. According to Bhowmick et al., 2020 analysing the stock market plays an important role in modern economic and financial activities for any country or region [2]. The uncertain risk in the stock market is one of the important measures to observe the impact on the volatility of stock index return.

According to Liu et al., 2021, empirical studies of prediction in volatility modelling of oil prices in the financial market, where the investors want to know economic evaluations of forecasting performance [3]. If volatility modelling in crude oil stock prices can predict the prices of other global stock variables whose price is based on a change in the prices of crude oil then the implications of oil volatility for financial markets are important for the global financial market. According to Nti et al., 2020, stock market prediction uses technical tools or fundamental analysis with the help of various statistical tools [4]. The stock market prices always vary from time to time may be daily or hourly for any company. Therefore, calculating the stock market price of any company can be determined with the help of proper analytical tools and techniques.

According to Saha et al., 2022, stock market prediction is considered a major challenge in the financial market and it may lead to profits by taking proper strategy [5]. Therefore, the use of graph-based techniques to handle market volatility emerges as a straightforward approach. Individual stocks in the market are connected. Therefore, many investors analysed the stock market using graph-based methods to get accurate values to predict the stock market indices.

According to Lu et al., 2023, the performances of stock markets are time varying; thus, it is necessary to incorporate Markov regime-switching with the help of different time series analysis methods.

3. Methodology

Stock market analysis is done for deciphering economic trends, making informed decisions about investments, and tackling financial risks. It helps in reading the performance and health of the companies, sectors, as well as the economy. The data of S&P BSE BANKEX and S&P BSE Energy were provided by the S&P BSE website for analysing their relationship. In this study, S&P BSE Energy index was taken as an independent variable, whereas the S&P BSE BANKEX index was taken as a dependent variable. These two indices have been chosen for study with a view to understanding the influence of energy in the overall performance of the market place.

First, the descriptive statistics were drawn to compute inference about the measures of central tendency, dispersion, and the pattern of distribution of data. Thus, the mean, median, standard deviation, skewness, and kurtosis have been computed for both the indices. Pearson moment correlations have been drawn to assess the linear relationship that exists between S&P BSE BANKEX and S&P BSE Energy indices. The correlation coefficient thus obtained has computed the strength and direction of association. Thereafter, the relationship between the indices was modelled using linear regression analysis. The S&P BSE BANKEX index was forecasted for the values of the S&P BSE Energy index. The regression equation, the coefficient of determination that is R^2, and regression coefficients were computed. Regression Model significance testing has been performed to validate if there exists an association between the indices; assuming the null hypothesis of no relationship exists between the indices. The obtained p-value was interpreted about the evidence of the null hypothesis; if it was less than 0.05, then the results were considered to be statistically significant. Here, ANOVA was used for comparing the means and corroborating the accuracy of the regression model.

ARIMA is used to understand the nature of fitted data to obtain predictions of future values of the S&P BSE BANKEX index, which encompasses a rich variety of components, viz., trend and seasonality in data to yield more accurate predictions. Also, to be noted that parameters of the ARIMA model have been optimized. Diagnostic checks are also conducted to check the adequacy of the model. These results showed the predictions of the movements and directions in which the S&P BSE BANKEX index would move based on the S&P BSE Energy index. ARCH model has been used to analyse the volatility and time-varying variance in stock returns, helping to understand how volatility evolves over time. GARCH model is used to analyse the past volatility with over time, providing a more comprehensive approach to modelling volatility persistence and predicting future volatility. On the other hand, LSTM models are employed to capture complex, non-linear relationships between stocks by learning patterns from historical data, making them effective for forecasting price movements or identifying dependencies between stock returns. Together, these models help to analyse and predict stock behaviour from different perspectives—volatility, persistence, and non-linear dependencies.

4. Results and Decisions

This context will observe the descriptive statistics of the daily return of S&P BSE Energy and the daily return of S&P BSE BANKEX. All the data has been collected for the BSE website from 1st April 2016 to 31st January 2025. A total of 2192 days and data have been used. Descriptive statistics provides the basic information about the data which defines the whole dataset.

	Daily Return Bankex	Daily Return Energy
Mean	0.0006109	0.000780
Standard Deviation	0.01419871	0.01469416

Table 1: Descriptive Statistics

From the above table, TABLE 1, it has been observed that both the mean value and standard deviation value are low. And the value of Pearson's correlation coefficient value is 0.5494744. This means there is a positive association exists between the two stock indices.

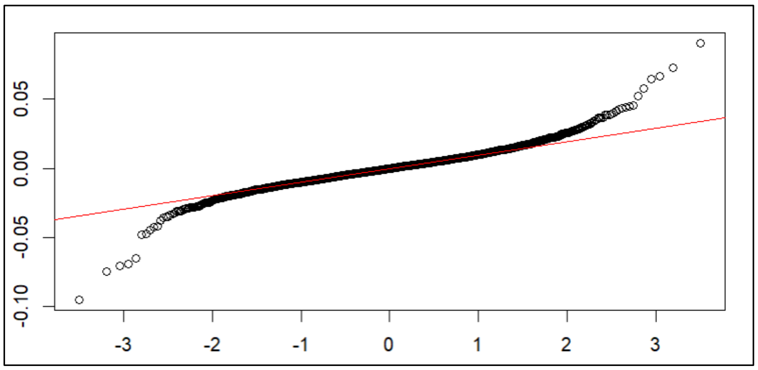

Figure 1: Normal Q-Q Plot

After observing Normal Q-Q plot from Fig-1, we have to examine the value of the dependent variable for individual stocks for other stocks to collect some information concerning the available explanatory variables, or to estimate the effect of some explanatory variable on the dependent variable linear regression has been used. Hence, contrast the alternative hypothesis H1: linearity exists with the null hypothesis H0: there is no linearity between them. S&P BSE BANKEX is regarded as the dependent variable, with S&P BSE Energy as the independent variable. From the analysis,

	Estimate	Standard Error	t-value	Pr(>|t|)
Intercept	0.0004330	0.0002626	1.648	0.0994
Daily Return Bankex	0.5686477	0.0184852	30.762	<2e-16

Table 2: Regression Analysis

In this model, the value of Multiple R-square is 0.6291, therefore, there is 62.91% variation within the stock markets, and the value of adjusted R-square is 0.6283 the model is 62.83% wealthy for our regression test. In our regression analysis, observing values from Table -2, the calculated p-value is less than 2e-16 which is less than 0.05. Therefore, there is sufficient evidence to reject the null hypothesis, which leads to the decision that there is linearity exists between the two stocks. To confirm our hypothesis, the residuals of our regression equation,

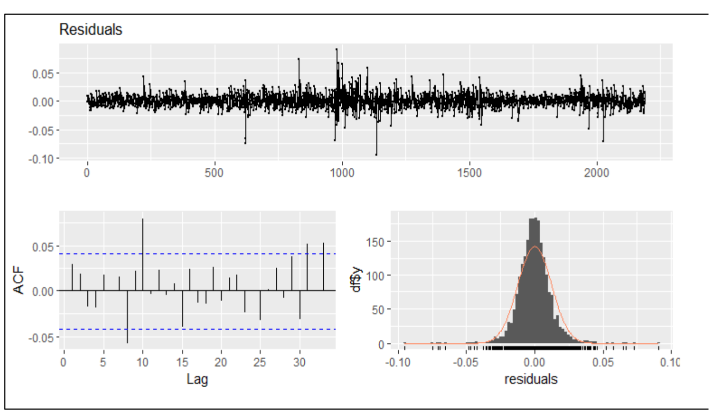

Figure 2: Residual Models

The outputs in FIG 2 are displaying the plots of the residuals against the fitted values of the model. The residuals are the differences between the actual response values and the predicted values from the model. The fitted values are the predicted values of the response variable based on the predictors in the model. In FIG-2 the residuals are normally distributed with a mean of zero and constant variance. therefore, the model does fit the data. To confirm the hypothesis testing, the ANOVA test has been used,

	Degrees of Freedom	Sum of Square	Mean Sum of Square	F-value	Pr(>|F)
Daily Return Energy	1	0.14270	0.142702	946.32	<2.2e-16
Residuals	2188	0.32994	0.000151		

Table 3: ANOVA Table

From the above TABLE 3, ANOVA modelling, the calculated p-value is less than 2.2e-16 which is less than 0.05 (p-value). Therefore, there is sufficient evidence to reject the null hypothesis, and it concludes that there is linearity exists between S&P BSE Energy and S&P BSE BANKEX. Now, to forecasting taken stocks for a given period, and to determine future demand for in investing those stocks, ARIMA (0,0,0) model with zero means has been used. The selection of ARIMA (0,0,0) as the appropriate model for a particular time series data depends on the characteristics of the data such as seasonality, trends, and the presence of autocorrelation.

In this context, ARIMA models are commonly used to predict the future price of the stock market based on previous prices of that stock market. ARIMA (0,0,0) with zero means gives a log-likelihood value of 6529.06, and an AIC value of -16056.12 therefore, the model is a better fit than other models and BIC value of -13050.43, therefore, the model has the largest modulus and preferred to be a fit model than others.

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training Set	5.01288e-20	0.01227432	0.00870461	74.87319	253.1958	0.8649015	0.03082175

Table 4: Training Set Error Measure

From the Table-4, the calculated values of residuals of our ARIMA (0,0,0) with zero mean, the Ljung-Box test gives the p-value as 0.001467.

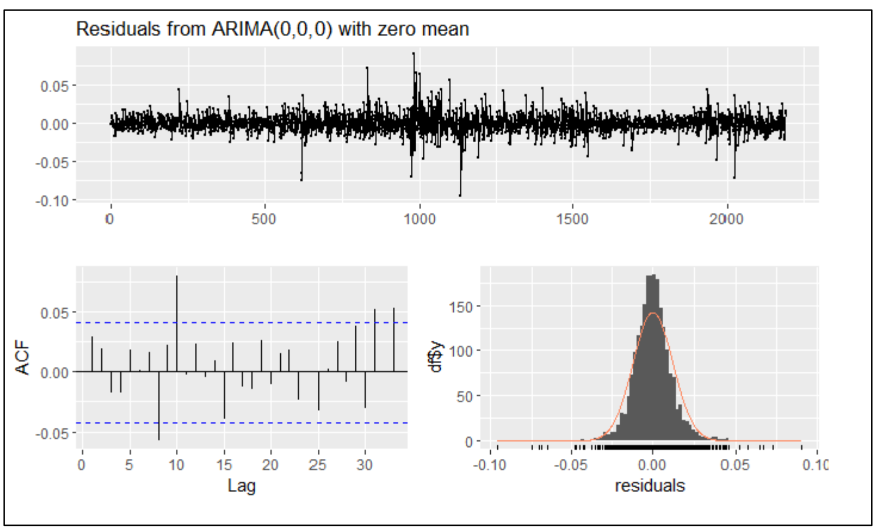

Figure 3: Residuals from ARIMA Model

The plots of the residuals against the fitted values are shown in the output of FIG. 3. The discrepancies between the actual response values and the model's anticipated values are known as residuals. The response variable's fitted values are the values that are expected based on the model's predictors. The residuals in FIG. 3 have a normal distribution with a mean of 0 and constant variance. Therefore, based on the calculated p-value, it can be concluded that there is no significant evidence of autocorrelation in the time series at the given level of significance. Therefore, the findings lead to the decision that there is a positive association between S&P BSE Energy and S&P BSE BANKEX.

Mean model					
	Coefficient	**σ**	**T**	**Pr(>|t|)**	**95% C.I.**
Mean (μ)	9.9974e-02	2.407e-04	4.153e-02	0.967	[-4.618e-04 ,4.818e-04]
Volatility model					
	Coefficient	**σ**	**T**	**Prp(>|t|)**	**95% C.I.**
ω (constant)	8.9082e	5.757e-06	15.475	5.142e-54	[7.780e-05, 1.004e-04]
α_1(ARCH coeff)	0.3660	6.436e-02	5.687	1.291e08	[0.240, 0.492]

Table 5: ARCH Model

In Table-5, the ARCH model states the constant mean of BSE Bankex returns is nearly zero, suggesting that the average return is insignificant over the observed period. Whereas, the volatility model, shows a significant base level of volatility (ω), indicating inherent fluctuations in BSE Bankex returns. The key ARCH coefficient ($\alpha 1$) is 0.3660, meaning past volatility has a moderate but notable effect on current volatility, with high volatility periods tending to persist. This result is statistically significant, highlighting the clustering nature of volatility. Therefore, this analysis illustrates how past volatility plays a crucial role in shaping current volatility in BSE Bankex.

Mean model					
	Coefficient	**σ**	**T**	**Pr(>|t|)**	**95% C.I.**
Mean (μ)	1.2265e-03	4.616e-04	2.657	7.883e-03	[3.218e-04, 2.131e-03]
Volatility model					
	Coefficient	**σ**	**t**	**Pr(>|t|)**	**95% C.I.**
ω (constant)	5.7845e-05	6.904e-06	8.378	5.363e-17	[4.431e-05, 7.138e-05]
α_1(ARCH coeff)	0.400	0.126	1.581	0.114	[-4.791e-02, 0.448]
β_1 (GARCH Coeff)	0.500	0.106	2.824	4.748e-03	[9.176e-02, 0.508]

Table 6: GARCH Model

In the above Table-6 n this GARCH model, the average return of BSE Bankex is statistically different from zero, with a p-value of 0.0078 for the mean coefficient (mu). There is strong evidence of volatility persistence, primarily driven by the significant GARCH coefficient ($\beta 1$), indicating that past volatility plays a significant role in predicting current volatility.

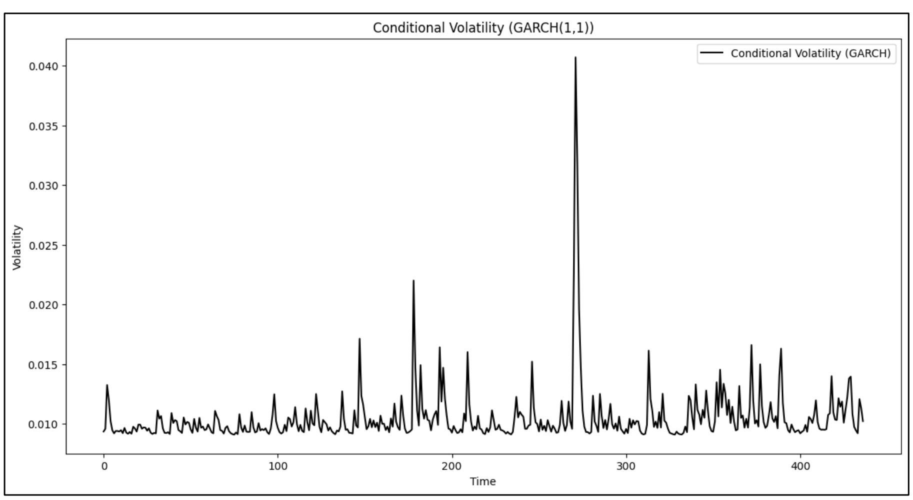

Figure 4: GARCH (1,1) Model

The above GARCH model shows the estimated conditional volatility over time, with volatility typically low but fluctuating over time to time. In GARCH (1,1) model, taken lag periods of high volatility tend to cluster together, while low volatility persists. Comparing the GARCH model's performance with other models and conducting residual tests are essential steps for validating the model's assumptions and fit. Overall, the graph highlights typical volatility patterns in financial time series, with the GARCH (1,1) model providing a solid framework for understanding these fluctuations.

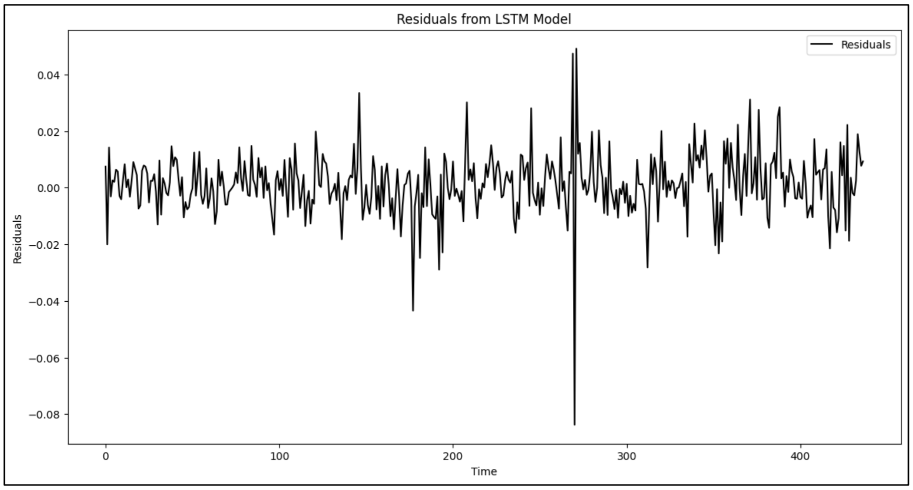

Figure 5: LSTM Model

From the above figure, the residual plot indicates that the LSTM model is performing well, capturing the underlying patterns in the data with reasonable accuracy. Analysed LSTM model also indicates the residuals fluctuate around zero, suggesting that the model's predictions are generally accurate on average. However, the relatively small magnitude of the residuals which is ranging from approximately -0.04 to +0.04, implies that the model is effectively capturing the key patterns.

5. Conclusion

Predicting stock prices is a challenging and risky endeavour due to the complexity of financial markets and the influence of investor perceptions and emotions. Stock prices often rise as investors believe they will continue to do so, creating a self-fulfilling prophecy. To navigate the stock market successfully, investors must exercise caution, conduct thorough analysis, and acknowledge its inherent unpredictability. Avoiding losses entirely is impossible, making it crucial to approach stock investment with a balanced understanding of the risks involved. This paper aimed to establish a meaningful relationship between S&P BSE Energy and S&P BSE BANKEX by employing an appropriate scaling method and examining the residual plot. The spikes in volatility observed in the residuals point to potential heteroscedasticity, indicating that while the LSTM model captures overall patterns, there may still be periods of non-constant variance in the data, which could be explored further using a GARCH model. By delving into the intricate dynamics of these stocks, researchers can contribute to the advancement of financial knowledge and enable investors to make more informed choices.

The established linearity offers a solid foundation for future investigations, fostering a deeper understanding of the complex interplay between these two stocks in the dynamic financial market. Sometime volatility in the financial market indicates sharp price fluctuations, often driven by other factors such as market events, company news, sentiment shifts, or technical factors. Despite fluctuations, volatility reverts to a baseline level, typical of financial markets In conclusion, this study successfully demonstrated the existence of a significant positive association between S&P BSE Energy and S&P BSE BANKEX. The robust correlation opens up avenues for advanced analytical approaches, benefiting researchers and investors alike. Through these endeavours, the financial landscape can be better navigated, leading to more prudent investment strategies and risk management techniques [6-12].

Further Scope of The Research

This paper consists of 992 days of data and forecasts them to find the association between two stock markets. Researchers can go for further statistical modelling by using various statistical tools and techniques.

Limitations of The Study

Analysing stock markets might be challenging but predicting the stock market is a foolish game for traders and investors for a shorter period. Predicting the stock market for a larger period may lead to profitability for investors. But it depends on Analysts to guess the proper strategy and use proper statistical tools to examine the nature and predict the future outcome.

Ethical Standards

This stock market analysis was conducted with unwavering adherence to ethical principles. By maintaining the highest standards of research integrity, valuable insights were aimed to be contributed to the field of financial analysis, and the trust of the readers and the wider research community was upheld.

Fundings

This work did not receive any fundings for support

Declaration of Interest

No conflict of interest is declared by the author

Data Sharing Statement

Data is available in BSE official website.

References

1. Yang, R., Yu, L., Zhao, Y., Yu, H., Xu, G., Wu, Y., & Liu, Z. (2020). Big data analytics for financial Market volatility forecast based on support vector machine. *International Journal of Information Management, 50*, 452-462.
2. Bhowmik, R., & Wang, S. (2020). Stock market volatility and return analysis: A systematic literature review. *Entropy, 22*(5), 522.
3. Liu, L., Geng, Q., Zhang, Y., & Wang, Y. (2022). Investors' perspective on forecasting crude oil return volatility: Where do we stand today?. *Journal of Management Science and Engineering, 7*(3), 423-438.
4. Nti, I. K., Adekoya, A. F., & Weyori, B. A. (2020). A systematic review of fundamental and technical analysis of stock market predictions. *Artificial Intelligence Review, 53*(4), 3007-3057.
5. Saha, S., Gao, J., & Gerlach, R. (2022). A survey of the application of graph-based approaches in stock market analysis and prediction. *International Journal of Data Science and Analytics, 14*(1), 1-15.
6. Chatziantoniou, I., Filippidis, M., Filis, G., & Gabauer, D. (2021). A closer look into the global determinants of oil price volatility. *Energy Economics, 95,* 105092.
7. Choi, S. Y. (2022). Evidence from a multiple and partial wavelet analysis on the impact of geopolitical concerns on stock markets in North-East Asian countries. *Finance Research Letters, 46,* 102465.
8. Gandhmal, D. P., & Kumar, K. (2019). Systematic analysis and review of stock market prediction techniques. *Computer Science Review, 34,* 100190.
9. Shah, D., Isah, H., & Zulkernine, F. (2019). Stock market analysis: A review and taxonomy of prediction techniques. *International Journal of Financial Studies, 7*(2), 26.
10. Sisodia, P. S., Gupta, A., Kumar, Y., & Ameta, G. K. (2022, February). Stock market analysis and prediction for NIFTY50 using LSTM Deep Learning Approach. In 2022 2nd international conference on innovative practices in technology and management (ICIPTM) (Vol. 2, pp. 156-161). *IEEE.*
11. Yuan, D., Li, S., Li, R., & Zhang, F. (2022). Economic policy uncertainty, oil and stock markets in BRIC: Evidence from quantiles analysis. *Energy Economics, 110,* 105972.
12. Zeng, S., Jia, J., Su, B., Jiang, C., & Zeng, G. (2021). The volatility spillover effect of the European Union (EU) carbon financial market. *Journal of Cleaner Production, 282,* 124394.