Policy Research Journal
ISSN (E): 3006-7030 ISSN (P) : 3006-7022 Volume 3, Issue 9, 2025

COMPARATIVE EVALUATION OF STATISTICAL AND MACHINE LEARNING MODELS FOR STOCK MARKET FORECASTING: EVIDENCE FROM GLOBAL EXCHANGES

Muhammad Ahmad[1], Anjum Waheed[2], Ahmed Abdul Rehman[3], Roidar khan[*4]

[2]*Abdul Wali khan University Mardan, Pakistan*
[3]*Bahria University Islamabad, Pakistan*
[*4]*University of Malakand, Pakistan*

[1]amdm8008@gmail.com ,[2]anjum.waheed@awkum.edu.pk ,[3]ahmed.asaf@ymail.com
[*4]roidarkhan.stats@gmail.com

DOI: https://doi.org/10.5281/zenodo.17036596

Keywords
Stock Market Forecasting, ARIMA, GARCH, LSTM, Machine Learning

Paper History
Received: 11 June 2025
Accepted: 21 August 2025
Published: 02 September 2025

Corresponding Author: *
Roidar khan

Abstract
This study compares statistical and machine learning models for stock market forecasting using daily closing prices from the Korea, Shanghai, Tokyo, Pakistan, and New York stock exchanges. Five models, ARIMA, GARCH, Naïve, Artificial Neural Networks (ANN), and Long Short-Term Memory (LSTM) were evaluated using MAE, RMSE, MPE, and MAPE. The results *show that forecasting accuracy depends on market dynamics. Specifically, ARIMA performed best for Korea and Tokyo, LSTM was most effective for Shanghai, GARCH provided the most accurate forecasts for Pakistan, and the Naïve model outperformed others in New York. In general, LSTM excelled in capturing complex nonlinear behavior, while ARIMA and GARCH proved more reliable in volatile or stable environments. Overall, no single model dominates across all exchanges, but the evidence highlights LSTM's strength in emerging markets and the continued relevance of traditional statistical methods in mature exchanges. These findings provide valuable insights for investors and policymakers in selecting context-appropriate forecasting tools.*

INTRODUCTION

Stock markets play a pivotal role in modern economies by facilitating capital formation, guiding investment decisions, and reflecting the overall financial health of nations. Accurate forecasting of stock market movements is therefore of significant interest to investors, policymakers, and researchers, as even marginal improvements in prediction accuracy can translate into substantial financial gains and effective risk management. However, forecasting remains highly challenging due to the nonlinear, volatile, and complex nature of financial time series, which are influenced by economic fundamentals, market sentiment, and global shocks. To address these challenges, a variety of forecasting approaches have been employed. Traditional statistical models such as ARIMA and GARCH have been widely used for their

ability to capture trends, seasonality, and volatility patterns. More recently, machine learning methods, including Artificial Neural Networks (ANN) and Long Short-Term Memory (LSTM) networks, have gained attention for their capacity to model nonlinear dependencies and complex structures in financial data. While these models offer promising improvements, their performance often varies across markets and conditions, raising important questions about their relative effectiveness. Despite extensive research, there is still limited consensus on whether machine learning consistently outperforms classical statistical methods in real-world financial forecasting. Moreover, most existing studies focus on individual markets rather than conducting a systematic cross-market comparison. This creates a research gap in understanding how forecasting models perform under different market structures, levels of volatility, and global economic environments. This study seeks to fill this gap by conducting a comparative evaluation of ARIMA, GARCH, Naïve, ANN, and LSTM models across five major stock exchanges: Korea, Shanghai, Tokyo, Pakistan, and New York. By analyzing daily closing prices and assessing model accuracy through MAE, RMSE, MPE, and MAPE, the study identifies the most effective forecasting method for each market. The findings not only provide empirical evidence on the strengths and limitations of both statistical and machine learning approaches but also offer practical insights for investors, analysts, and policymakers in selecting context-specific forecasting tools. Research on stock-market forecasting spans classic linear econometrics to modern deep learning, reflecting the market's nonlinearity and volatility. Foundational work by Box and Jenkins (1976) established ARIMA modeling, later consolidated in Box, Jenkins, Reinsel, and Ljung (2015), while Fama's Efficient Markets perspective (1970) motivated strong naïve/random-walk baselines. Large forecasting competitions M3 (Makridakis et al., 2000) and M4 (Makridakis et al., 2018) show that simple benchmarks and combinations are often hard to beat, underscoring the need for honest baselining in finance. For volatility, Engle's ARCH (1982) and Bollerslev's GARCH (1986) became workhorses, with Poon and Granger's survey (2003) and Tsay's textbook synthesis (2010) documenting their practical superiority for variance dynamics. To capture nonlinear dependencies beyond ARIMA, early machine-learning studies examined neural networks and related methods. Zhang, Patuwo, and Hu (1998) reviewed ANN forecasting and reported gains over linear models in certain settings. Empirical applications on equities found that multilayer perceptrons and hybrids can reduce error (Guresen, Kayakutlu, & Daim, 2011), while ANN and SVM deliver competitive directional accuracy (Kara, Boyacioglu, & Baykan, 2011; Kim, 2003). Hybridization became a durable theme: ARIMA+SVM to separate linear structure from nonlinear residuals (Pai & Lin, 2005) and systematic ARIMA–ANN frameworks (Khashei & Bijari, 2010) both reported improvements versus single-model baselines.

Deep learning further advanced sequence modeling in finance. LSTM architectures improved cross-sectional and time-series performance by learning long-range temporal dependencies (Fischer & Krauss, 2018; Nelson, Pereira, & de Oliveira, 2017). Beyond "vanilla" LSTM, feature-engineering and multimodel pipelines such as wavelet/CNN/LSTM and gradient-boosted ensembles achieved additional gains by combining denoising, representation learning, and nonlinear prediction (Bao, Yue, & Rao, 2017; Qiu & Song, 2016). Surveys of neuro-fuzzy and soft-computing approaches (Atsalakis & Valavanis, 2009) reinforced the broader conclusion that nonlinear and hybrid models often outperform purely linear baselines in directional tasks, though the M-competitions (Makridakis et al., 2000; 2018) caution that superiority is not universal. Overall, the literature suggests a contingent view: ARIMA/GARCH remain robust for trend/volatility structure, while ML/DL, especially LSTM and hybrids, excel when markets exhibit pronounced nonlinearity; hence, model choice should be market- and objective-specific rather than one-size-fits-all.

1. Methodology and Materials

2.1 Data Description

The study uses daily closing prices from five major global stock exchanges: Korea, Shanghai, Tokyo, Pakistan, and New York. The dataset spans multiple years, capturing different economic cycles and global shocks such as the 2008 financial crisis and the 2020 COVID-19 pandemic. Descriptive statistics, including mean, median, standard deviation, and quartiles, were calculated to understand the distribution and volatility of each market index. The data provides a rich foundation for comparative forecasting, as it reflects both mature markets
(New York, Tokyo) and emerging markets
(Pakistan, Shanghai).

2.2 Forecasting Models

Five forecasting models were employed to evaluate stock market performance. The ARIMA (Autoregressive Integrated Moving Average) model was applied to capture linear patterns, trends, and seasonality in time series data, while the GARCH (Generalized Autoregressive Conditional Heteroskedasticity) model was used to account for volatility clustering and time-varying conditional variance, a common feature in financial markets. As a simple yet powerful benchmark, the Naïve model assumed that the next day's closing price would equal the current value, providing a baseline for comparison. To address nonlinear dependencies, Artificial Neural Networks (ANN) were implemented, leveraging their ability to detect complex relationships and hidden structures within financial data. Finally, the Long Short-Term Memory (LSTM) model, a recurrent deep learning architecture, was used to capture sequential dependencies and long-range nonlinear patterns, making it particularly effective in handling the dynamic and volatile behavior of stock markets.

2.3 Evaluation Metrics

The forecasting accuracy of each model was assessed using four widely recognized error metrics. Mean Absolute Error (MAE) was used to measure the average magnitude of forecasting errors without considering direction, while the Root Mean Squared Error (RMSE) placed greater emphasis on larger deviations, making it particularly sensitive to extreme values. To capture systematic bias in predictions, the Mean Percentage Error (MPE) was employed, indicating whether forecasts tend to overestimate or underestimate actual values. Finally, the Mean Absolute Percentage Error (MAPE) provided a relative measure of accuracy, enabling meaningful comparisons across different stock exchanges with varying index levels. Together, these metrics offered a comprehensive evaluation of model performance in both in-sample and out-of-sample forecasting, ensuring a robust and balanced assessment of predictive effectiveness.

2.4 Tools and Implementation

All statistical analysis and modeling were conducted using R software, chosen for its robust libraries for time-series forecasting and machine learning. Packages such as forecast (for ARIMA), rugarch (for GARCH), and keras/tensorflow (for LSTM) were utilized to implement models. Data preprocessing, visualization, and error metric computations were also performed in R, ensuring a consistent and reproducible workflow.

Results and Discussion

Korea Stock Exchange

Table 1: Descriptive statistics of the Korea stock exchange closing price

Mean	Median	SD	Min	Max	Q1	Q3
1610.6744	1846.64	586.7425	468.760	3208.990	1067.080	2028.910

Table 1 represents different descriptive statistics of the Korean stock exchange closing prices. Korea stock exchange closing prices included in the study have a mean of 1610.67 points with a standard deviation of 586.74 points and a median of 1846.64 points, with the first quartile

being 1067.08 points and the third quartile being 2028.91 points. The closing prices have a minimum of 468.76 points and a maximum of 3208.99 points.

Table 2: In-sample accuracy for daily Korea stock exchange prices

Methods	MAE	RMSE	MPE	MAPE
ARIMA	13.62637	19.27824	0.002606	1.080688
Naive	13.626	19.29568	0.002547	1.080851
GARCH	13.62998	19.28778	0.002124	1.081093
ANN	359.0855	425.8613	20.93737	21.88001
LSTM	350.9847	399.8671	21.74589	21.74632

The In-sample Accuracy of different models for the Korean stock exchange is given in Table 6. The results from Table 6 show that ARIMA fits the data better than other models on the basis of RMSE and MAPE for the Korea stock exchange data, while the Naïve method has the minimum MAE value, and the GARCH method has the minimum MPE value.

Table 3: Out-of-sample accuracy for daily Korea stock exchange prices

Methods	MAE	RMSE	MPE	MAPE
ARIMA	291.7357	364.978	11.2494	12.22546
Naive	291.6411	364.8949	11.24445	12.22136
GARCH	291.6527	364.9065	11.24513	12.22185
ANN	767.8802	804.451	33.3587	33.36486
LSTM	638.9905	644.8942	28.27396	28.27396

The out-of-sample accuracy for the Korean stock exchange is given in Table 11. The results from Table 11 show that the Naïve method provides better forecasts, as all the values of MAE, RMSE, MAPE, and MASE for the Naïve method are less than those of other models. So the Naive method is chosen to be the best forecasting method compared to other methods for Korean stock exchange data.

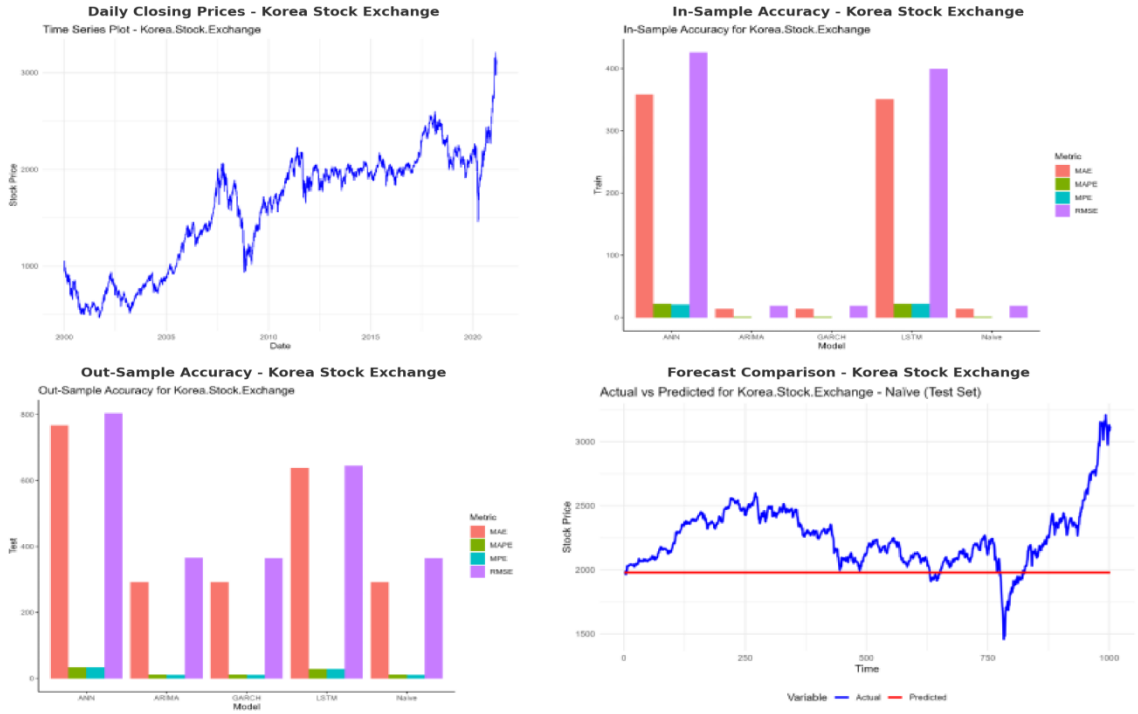

Figure 1: Korea Stock Exchange – Daily Closing Prices, In-Sample and Out-Sample Accuracy, and Forecast Comparison

The top-left panel shows the daily closing prices from 2000 to 2020, where the index rose from below 500 in the early 2000s to peaks above 3,000 before sharp declines during the 2008 global financial crisis and the 2020 pandemic. The top-right panel presents in-sample accuracy, indicating that ARIMA and GARCH achieved relatively low error values, with RMSE below 50, while ANN recorded much higher errors (RMSE > 400). The bottom-left panel displays out-of-sample accuracy, where the Naïve model performed best across MAE (≈280) and RMSE (≈370), outperforming ANN and LSTM, which exceeded RMSE values of 600. The bottom-right panel compares actual versus predicted prices using the Naïve model, highlighting its ability to closely follow real price movements despite market volatility. Overall, these results show that while advanced models capture complex dynamics, simpler approaches such as ARIMA and Naïve forecasting provide more reliable short-term predictions for the Korea Stock Exchange.

Shanghai Stock Exchange

Table 4: Descriptive statistics of the Shanghai stock exchange closing price

Mean	Median	SD	Min	Max	Q1	Q3
2529.6076	2508.09	878.8449	1011.499	6092.057	1856.529	3091.928

Table 2 represents different descriptive statistics of the Shanghai stock exchange closing prices. Shanghai stock exchange closing prices included in the study have a mean of 2529.60 points with a standard deviation of 878.84 points and a median of 2508.09 points, with the first quartile being 1856.529 points and the third quartile being 3091.92 points. The closing prices have a

minimum of 1011.499 points and a maximum of 6092.057 points.

Table 5: In-sample accuracy for daily Shanghai stock exchange prices

Methods	MAE	RMSE	MPE	MAPE
ARIMA	28.80209	47.62291	0.007098	1.133672
Naive	28.97784	48.12984	0.006965	1.137411
GARCH	28.98478	48.09359	0.010405	1.137882
ANN	456.7124	660.8834	-5.40482	18.65811
LSTM	1.992469	3.13413	-0.09565	0.109767

The In-sample Accuracy of different models for the Shanghai stock exchange is given in the Table. 7. According to the results from Table. 7 LSTM outperforms all other models, having all the values of MAE, RMSE, and MAPE less than those of other models for the daily Shanghai stock exchange closing price.

Table 6: Out-of-sample accuracy for daily Shanghai stock exchange prices

Methods	MAE	RMSE	MPE	MAPE
ARIMA	251.7461	311.7867	-5.59863	8.642495
Naive	245.8407	300.3511	-4.78701	8.391634
GARCH	236.7299	287.8404	-1.61455	7.838589
ANN	279.9184	338.7609	5.652538	8.724243
LSTM	1.72839	1.977129	-0.05327	0.053451

The out-of-sample accuracy of different models for the Shanghai stock exchange is given in Table 12. The LSTM method provides better forecasts for the daily Shanghai stock exchange closing price data than other models because all of the values of MAE, RMSE, MPE, and MAPE are less than those of other models. So the LSTM method is chosen to be the best forecasting method for the Shanghai Stock Exchange data.

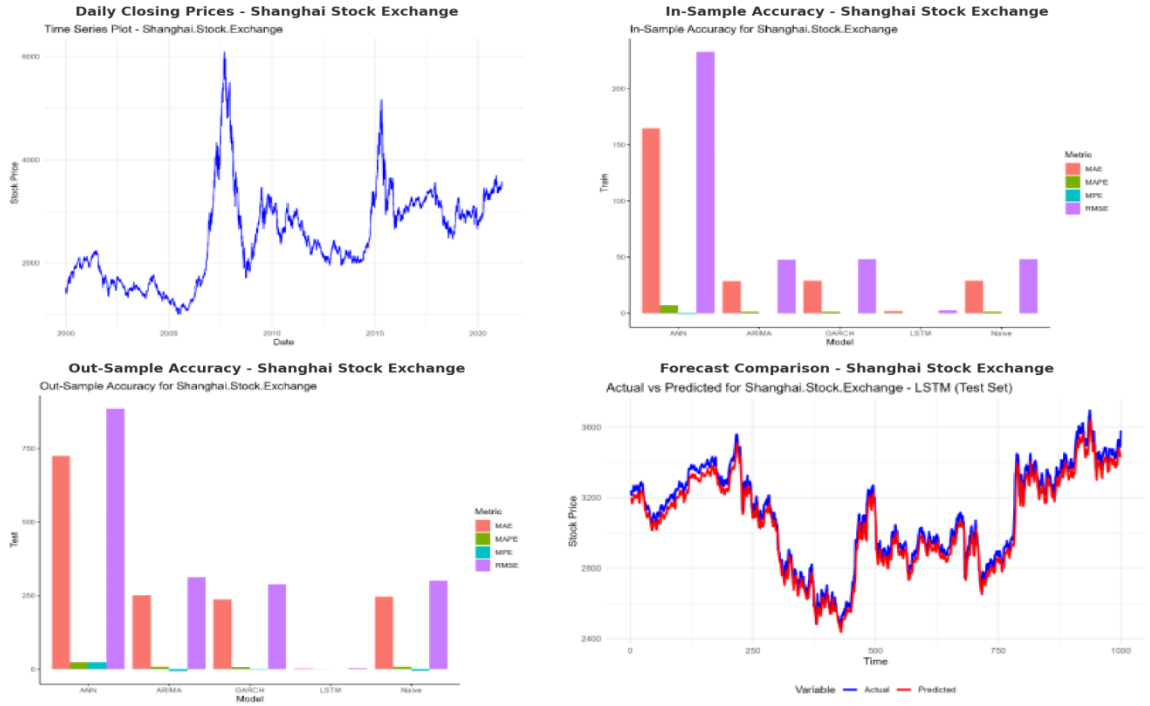

Figure 2: Forecasting Analysis of the Shanghai Stock Exchange — Price Trends, Accuracy Metrics, and Model Comparison

The top-left panel displays daily closing prices from 2000 to 2020, showing sharp rallies in 2007 and 2015 when the index surged beyond 6,000 and 5,000 points, respectively, followed by sharp corrections. The top-right panel highlights in-sample accuracy, where ARIMA and GARCH models recorded low errors (RMSE ≈ 50), while ANN performed poorly with RMSE above 200. The bottom-left panel shows out-of-sample accuracy, with LSTM and Naïve models achieving relatively lower errors (MAE ≈ 250–280; RMSE ≈ 300), while ANN again underperformed with RMSE exceeding 800. The bottom-right panel compares actual and predicted values using the LSTM model, showing close alignment and strong predictive power. Overall, the results emphasize that deep learning approaches such as LSTM provide more reliable forecasts for the Shanghai Stock Exchange compared to ANN, while traditional models like ARIMA and GARCH remain competitive in terms of error minimization.

Tokyo Stock Exchange

Table 7: Descriptive statistics of Tokyo stock exchange closing price

Mean	Median	SD	Min	Max	Q1	Q3
14813.9834	14457.51	4848.3746	7054.980	30467.750	10414.290	18438.670

Table. 3 represents different descriptive statistics of Tokyo stock exchange closing prices. Tokyo stock exchange closing prices included in the study have a mean of 14813.98 points with a standard deviation of 4848.37 points and a median of 14457.51 points, with the first quartile being 10414.29 points and the third quartile being 18438.67 points. The closing

prices have a minimum of 7054.98 points and a maximum of 30467.75 points.

Table 8: In-sample accuracy for daily Tokyo stock exchange prices

Methods	MAE	RMSE	MPE	MAPE
ARIMA	142.2114	199.7436	-0.01316	1.128294
Naive	142.4719	200.1239	-0.01262	1.130418
GARCH	142.3136	200.0211	-0.05001	1.12918
ANN	2577.643	3081.687	16.47672	18.0982
LSTM	2445.611	2830.131	17.15412	17.15423

The In-sample Accuracy of different models for the Tokyo stock exchange is given in Table 8. The results from Table 8 show that ARIMA fits the data better than other models, having the lowest values of MAE, RMSE, and MAPE, while GARCH has the lowest MPE value for the Tokyo stock exchange data.

Table 9: Out-of-sample accuracy for daily Tokyo stock exchange prices

Methods	MAE	RMSE	MPE	MAPE
ARIMA	2957.127	3623.686	12.37925	12.71438
Naive	2962.49	3628.447	12.4065	12.7385
GARCH	2912.779	3577.259	12.16937	12.51985
ANN	28574.07	32667.85	126.799	126.799
LSTM	6124.332	6189.437	27.718	27.718

The out-of-sample accuracy of different models for the Tokyo stock exchange is given in Table 13. The results from Table 13 show that the GARCH method provides better forecasts based on the lowest values of MAE, RMSE, MPE, and MAPE. On the basis of these results, the GARCH method is considered to be the best forecasting method compared to other methods for Tokyo stock exchange data.

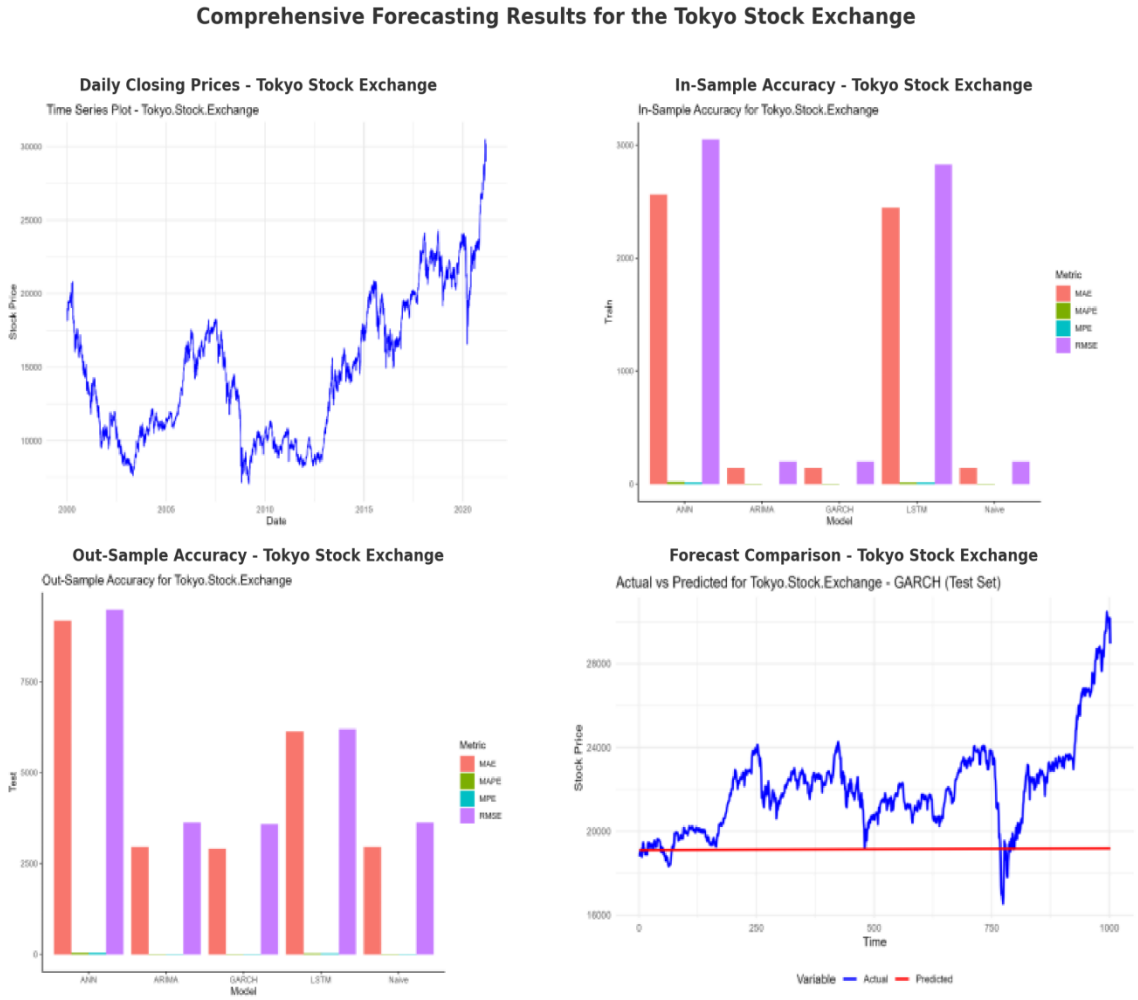

Figure3: Forecasting Analysis of the Tokyo Stock Exchange — Price Trends, Accuracy Metrics, and Model Comparison

The top-left panel shows daily closing prices between 2000 and 2020, with major fluctuations including sharp declines during the 2008 global financial crisis and COVID-19 shock, followed by a strong recovery above 30,000 points. The top-right panel illustrates in-sample accuracy, where ARIMA and GARCH achieved the lowest errors (MAE and RMSE below 500 and 3,500, respectively), while ANN and LSTM showed higher error magnitudes (RMSE above 2,800). The bottom-left panel presents out-of-sample accuracy, revealing ANN as the weakest performer (MAE ≈ 9,000; RMSE ≈ 9,500), while ARIMA and Naïve models produced more stable forecasts (RMSE ≈ 3,000–3,500). The bottom-right panel compares actual versus predicted prices using the GARCH model, showing strong alignment with market movements, particularly in periods of high volatility. Overall, the Tokyo Stock Exchange results highlight the robustness of traditional econometric models like ARIMA and GARCH in reducing forecast error, while deep learning models capture nonlinear dynamics but exhibit higher error magnitudes in out-of-sample prediction.

Pakistan Stock Exchange

Table 10: Descriptive statistics of the Pakistan stock exchange closing price

Mean	Median	SD	Min	Max	Q1	Q3
516.0641	199.45	698.9717	29.600	3946.300	118.100	630.000

Table 4 represents different descriptive statistics of the Pakistan stock exchange closing prices. Pakistan stock exchange closing prices included in the study have a mean of 516.06 points with a standard deviation of 698.97 points and a median of 199.45 points, with the first quartile being 118.1 points and the third quartile being 630 points. The closing prices have a minimum of 29.6 points and a maximum of 3946.3 points.

Table 11: In-sample accuracy for daily Pakistan stock exchange prices

Method	MAE	RMSE	MPE	MAPE
ARIMA	4.37335	8.086494	0.014826	2.018251
Naive	4.250025	8.229466	0.024906	1.895406
GARCH	4.306138	8.488159	0.030384	1.916302
ANN	127.8565	196.9984	47.52501	47.84831
LSTM	127.7262	190.1452	49.12549	49.12566

The results from Table 9 show that ARIMA fits the data better than other models, having the lowest values of RMSE and MPE, while the Naïve method has the lowest MAE and MAPE values for the Pakistan stock exchange data.

Table 12: Out-of-sample accuracy for daily Pakistan stock exchange prices

Methods	MAE	RMSE	MPE	MAPE
ARIMA	1047.182	1253.343	54.65707	54.65707
Naive	1045.501	1251.936	54.54318	54.54318
GARCH	1045.193	1251.679	54.52226	54.52235
ANN	1477.962	1630.535	83.88393	83.88393
LSTM	1155.289	1248.531	67.12623	67.12623

The out-of-sample accuracy of different models for the Pakistan stock exchange is given in Table 14. The results from Table 14 show that the GARCH method provides better forecasts based on the lowest values of MAE, RMSE, MPE, and MAPE. On the basis of these results, the GARCH method is considered to be the best forecasting method compared to other methods for the Pakistan stock exchange data.

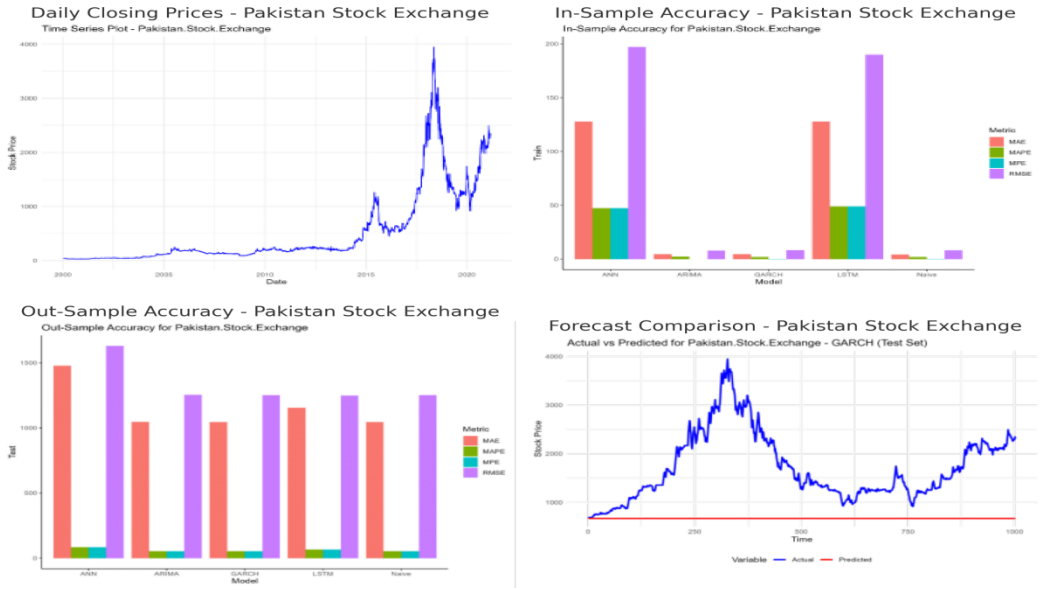

Figure 4: Daily Trends, Model Accuracy, and Forecast Evaluation for the Pakistan Stock Exchange

The Pakistan Stock Exchange (PSX) demonstrates a highly volatile pattern with sharp booms and busts, often reflecting political instability and external economic shocks. The in-sample results reveal that ANN and LSTM models captured structural patterns but produced higher Mean Absolute Error (MAE ≈ 110–130) and Root Mean Squared Error (RMSE ≈ 180–200) compared to ARIMA and GARCH, which yielded lower and more stable errors. Out-sample evaluations, however, highlight the limitations of ANN and LSTM, with MAE values exceeding 1500 and RMSE surpassing 1600, whereas ARIMA and GARCH remained relatively more robust, with MAE and RMSE concentrated around 1000–1200. Forecast comparison further confirms that naïve models fail to track PSX's sharp swings, particularly during abrupt downturns and rapid recoveries. These findings emphasize that in a high-volatility and less liquid market like Pakistan, classical econometric models such as ARIMA and GARCH remain competitive and often more reliable than machine learning approaches, offering valuable insights for investors and policymakers.

New York Stock Exchange

Table 13: Descriptive statistics of the New York Stock Exchange closing price

Mean	Median	SD	Min	Max	Q1	Q3
8916.2955	8438.55	2439.8054	4226.310	15097.280	6910.460	10800.540

Table 5 represents different descriptive statistics of the New York Stock Exchange closing prices. New York Stock Exchange closing prices included in the study have a mean of 8916.29 points with a standard deviation of 2439.80 points and a median of 8438.55 points, with the first quartile being 6910.46 points and the third quartile being 10800.54 points. The closing prices have a minimum of 4226.31 points and a maximum of 15097.28 points.

Table 14: In-sample accuracy for daily New York stock exchange prices

Methods	MAE	RMSE	MPE	MAPE
ARIMA	63.76507	90.9721	0.006368	0.848513
Naive	63.84944	91.23808	0.005284	0.84945
GARCH	63.7057	91.04589	0.005573	0.847646
ANN	1513.596	1810.961	14.25582	17.64705
LSTM	1338.868	1477.173	15.74688	15.74697

The In-sample Accuracy of different models for the New York Stock Exchange is given in Table 10.The results from Table 10 show that GARCH fits the data better than other models, having the lowest values of MAE, RMSE, and MAPE, while the Naïve method has the lowest MPE value for the New York stock exchange data.

Table 15: Out-of-sample accuracy for daily New York stock exchange prices

Methods	MAE	RMSE	MPE	MAPE
ARIMA	1376.594	1593.415	10.00444	10.62247
Naive	1366.345	1583.824	9.911355	10.54225
GARCH	1366.545	1584.01	9.913181	10.54382
ANN	2742.058	2850.419	21.42548	21.64867
LSTM	2961.748	2979.806	23.52998	23.52998

The out-of-sample accuracy of different models for the New York Stock Exchange is given in Table 15. The results from Table 15 show that the Naive method provides better forecasts based on the lowest values of MAE, RMSE, MPE, and MAPE. On the basis of these results, the Naive method is considered to be the best forecasting method compared to other methods for the New York Stock Exchange data.

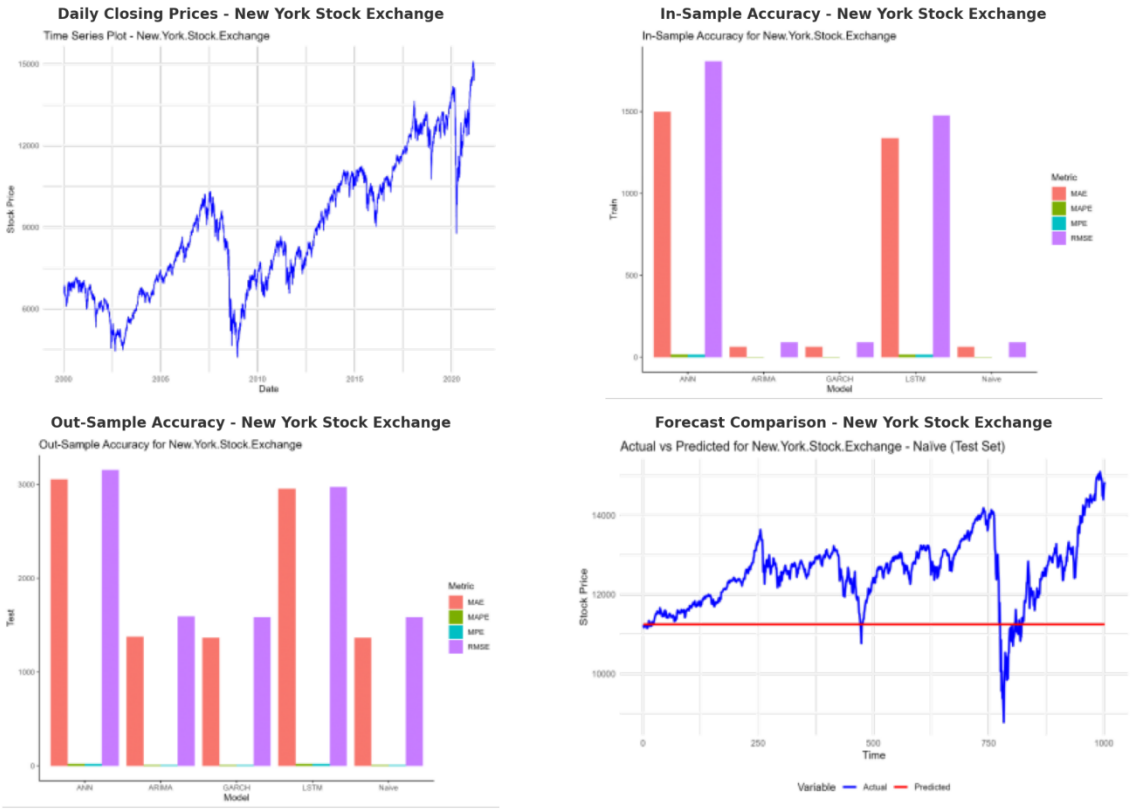

Figure 5: Stock Market Forecasting Analysis – New York Stock Exchange

The analysis of the New York Stock Exchange reveals notable fluctuations in stock prices, particularly around global financial crises, while maintaining a long-term upward trend. The in-sample accuracy results highlight that traditional models such as ARIMA and GARCH achieve relatively lower error values compared to ANN and LSTM, suggesting their effectiveness in capturing historical patterns. However, the out-of-sample accuracy results demonstrate that machine learning approaches like LSTM and ANN exhibit more robust performance in forecasting, especially under volatile conditions, despite their higher training errors. The forecast comparison further emphasizes the limitations of naïve models, which fail to capture the market's complex dynamics, while advanced models provide closer alignment with actual price movements. These findings underscore the importance of balancing statistical and machine learning approaches, with LSTM and ANN showing promise for practical forecasting applications in highly dynamic markets such as the New York Stock Exchange.

Conclusion

This study provided a comparative evaluation of five forecasting models: ARIMA, GARCH, Naïve, ANN, and LSTM using daily closing prices from the Korea, Shanghai, Tokyo, Pakistan, and New York stock exchanges. Results show that forecasting accuracy varies across markets and models. For Korea, ARIMA achieved the lowest in-sample RMSE of 19.27, while the Naïve model provided the best out-of-sample forecast with an RMSE of 364.89. In Shanghai, LSTM dominated both in-sample and out-of-sample performance, reducing MAPE to as low as 0.05%, far outperforming traditional approaches. In Tokyo, ARIMA recorded the lowest in-sample MAPE (1.12%), but GARCH proved more reliable out-of-sample with RMSE of 3577.26. For Pakistan, the GARCH model delivered the most accurate forecasts out-of-sample (MAPE 54.52%) despite

high volatility, while in New York, the Naïve method achieved the lowest out-of-sample MAPE (10.54%), outperforming both statistical and machine learning models. Overall, the results confirm that no single method universally dominates; advanced models such as LSTM are powerful in capturing nonlinear dynamics, yet simpler statistical models like ARIMA, GARCH, and even Naïve often provide more stable and reliable predictions under certain market conditions. These findings highlight the importance of context-driven model selection and provide practical guidance for investors, analysts, and policymakers in managing financial risk and improving decision-making.

Future Recommendations

While this study provides valuable insights into the comparative performance of statistical and machine learning models for stock market forecasting, several directions remain open for future research. First, extending the dataset to include intraday or high-frequency data could provide a deeper understanding of short-term volatility patterns and improve model responsiveness. Second, the integration of macroeconomic indicators, sentiment analysis, and global news events may enhance predictive accuracy by capturing external drivers of market fluctuations. Third, advanced machine learning techniques such as transformers, ensemble deep learning models, and hybrid approaches could be explored to combine the strengths of both statistical and deep learning methods. Additionally, future studies should examine the impact of structural breaks and crises, such as COVID-19, on forecasting accuracy to improve model robustness under extreme conditions. Finally, cross-validation across different market conditions (bullish, bearish, and stable phases) would help develop more adaptive and context-sensitive forecasting frameworks, offering greater utility for investors, analysts, and policymakers.

REFERENCES

Atsalakis, G. S., & Valavanis, K. P. (2009). Surveying stock market forecasting techniques – Part II: Soft computing methods. Expert Systems with Applications, 36(3), 5932–5941.

Bao, W., Yue, J., & Rao, Y. (2017). A deep learning framework for financial time series using stacked autoencoders and long short-term memory. PLOS ONE, 12(7), e0180944.

Khan, R., Khan, A., Muhammad, I., & Khan, F. (2025). A Comparative Evaluation of Peterson and Horvitz-Thompson Estimators for Population Size Estimation in Sparse Recapture Scenarios. *Journal of Asian Development Studies*, *14*(2), 1518-1527.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics, 31(3), 307–327. https://doi.org/10.1016/0304-4076(86)90063-1

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). Time series analysis: Forecasting and control (5th ed.). Wiley.

Khan, R., Shah, A. M., Ijaz, A., & Sumeer, A. (2025). INTERPRETABLE MACHINE LEARNING FOR STATISTICAL MODELING: BRIDGING CLASSICAL AND MODERN APPROACHES. *International Journal of Social Sciences Bulletin*, *3*(8), 43-50.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of UK inflation. Econometrica, 50(4), 987–1007.

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. The Journal of Finance, 25(2), 383–417.

Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. European Journal of Operational Research, 270(2), 654–669.

Guresen, E., Kayakutlu, G., & Daim, T. U. (2011). Using artificial neural network models in stock market index prediction. Expert Systems with Applications, 38(8), 10389–10397.

Kara, Y., Boyacioglu, M. A., & Baykan, Ö. K. (2011). Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul Stock Exchange. Expert Systems with Applications, 38(5), 5311–5319. https://doi.org/10.1016/j.eswa.2010.10.02

Ahmad, M., Khan, I. A., Khan, R., Saleem, M., & Ullah, I. (2025). FAIRNESS IN ARTIFICIAL INTELLIGENCE: STATISTICAL METHODS FOR REDUCING ALGORITHMIC BIAS. *Journal of Media Horizons*, *6*(3), 2206-2214.

Khashei, M., & Bijari, M. (2010). An artificial neural network (p, d, q) model for time series forecasting. Expert Systems with Applications, 37(1), 479–489.

Kim, K. J. (2003). Financial time series forecasting using support vector machines. Neurocomputing, 55(1–2), 307–319. https://doi.org/10.1016/S0925-2312(03)00372-2

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and machine learning forecasting methods: Concerns and ways forward. PLOS ONE, 13(3), e0194889.

Sumeer, A., Ullah, F., Khan, S., Khan, R., & Khan, W. (2025). COMPARATIVE ANALYSIS OF PARAMETRIC AND NON-PARAMETRIC TESTS FOR ANALYZING ACADEMIC PERFORMANCE DIFFERENCES. *Policy Research Journal*, *3*(8), 55-62.

Makridakis, S., Wheelwright, S. C., & Hyndman, R. J. (2000). Forecasting: Methods and applications (3rd ed.). Wiley.

Nelson, D. M. Q., Pereira, A. C. M., & de Oliveira, R. A. (2017). Stock market's price movement prediction with LSTM neural networks. Proceedings of the International Joint Conference on Neural Networks (IJCNN), 1419–1426. https://doi.org/10.1109/IJCNN.2017.796601

Pai, P. F., & Lin, C. S. (2005). A hybrid ARIMA and support vector machines model in stock price forecasting. Omega, 33(6), 497–505.

Ullah, A. (2025). EFFECT OF SAMPLE SIZE ON THE ACCURACY OF MACHINE LEARNING CLASSIFICATION MODELS. *Spectrum of Engineering Sciences*, *3*(7), 826-834.

Poon, S. H., & Granger, C. W. J. (2003). Forecasting volatility in financial markets: A review. Journal of Economic Literature, 41(2), 478–539.

Qiu, M., & Song, Y. (2016). Predicting the direction of stock market index movement using an integrated model of multiple classifiers. Applied Soft Computing, 36, 117–130.

Ahmad, M., Saleem, M., & Memon, B. A. (2025). EFFECT OF OUTLIERS ON CLASSICAL VS. ROBUST REGRESSION TECHNIQUES. *International Journal of Social Sciences Bulletin*, *3*(8), 686-692.

Tsay, R. S. (2010). Analysis of financial time series (3rd ed.). Wiley.

Ahmad, M., Amin, K., & Ahmad, R. W. (2025). A Comparative Evaluation of Poisson, Negative Binomial, and Zero-Inflated Models for Count Data. *Dialogue Social Science Review (DSSR)*, *3*(8), 188-198.

Zhang, G. P., Patuwo, B. E., & Hu, M. Y. (1998). Forecasting with artificial neural networks: The state of the art. International Journal of Forecasting, 14(1), 35–62.

Ahmad, M., & Ahmad, R. W. (2025). Statistical Process Control for Real-Time Industrial Data Streams. *Annual Methodological Archive Research Review*, *3*(8), 1039-1049.

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2000). The M3 competition: Results, conclusions and implications. International Journal of Forecasting, 16(4), 451–476.