WEP Warwick Evans Publishing

Transactions on Computer Science and Intelligent Systems Research
ISSN: 2960-1800, eISSN: 2960-2238 | Volume 10
ISET 2025

Financial Time Series Forecasting: A Hybrid Approach Combining AR-GARCH and Machine Learning Models

Shangrong Han*

Department of Statistical Science, University College London, London, UK

*Corresponding author: zcaksh4@ucl.ac.uk

Abstract. Accurately forecasting financial markets remains a central challenge in economics due to the volatile, nonlinear, and complex nature of asset price movements. Traditional statistical models like Autoregressive Generalized Autoregressive Conditional Heteroskedasticity (AR-GARCH) have been widely used for modeling volatility, but often fall short when addressing intricate market patterns. This study systematically compares the predictive capabilities of AR-GARCH, Long Short-Term Memory (LSTM) networks, Convolutional Neural Networks (CNN), Random Forest (RF), and a hybrid ensemble on weekly data from the S&P 500, FTSE 100, and Nikkei 225 indices spanning 2000 to 2024. Performance is evaluated using standard forecasting metrics: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). The findings indicate that while each separate model has varying strengths depending on market conditions, the hybrid model consistently achieves superior accuracy by leveraging the diversity of all approaches. These results highlight the benefits of integrating classical econometric approaches with contemporary machine learning techniques to improve forecasting precision and strengthen the robustness of models in financial time series analysis.

Keywords: Financial time series forecasting, Machine learning, Hybrid forecasting models, AR-GARCH, LSTM, CNN, Random Forest.

1. Introduction

Stock market prediction has long been a complex and widely studied topic in financial economics. Traditional time series models, such as the Autoregressive Conditional Heteroskedasticity (ARCH) model and its well-known extension, the Generalized ARCH (GARCH) model, have served as the theoretical foundation for volatility modeling for decades [1, 2]. While these models are praised for their statistical reliability and their ability to capture volatility clustering, they have clear limitations when it comes to handling nonlinear relationships and complex time dependencies [3]. As financial markets become more interconnected and influenced by multiple factors, these traditional models often fail to deliver sufficient forecasting accuracy [4].

The advancement and increasing adoption of modern machine learning techniques have brought major changes to financial modeling in recent years. Neural network architectures like Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) are increasingly used in financial forecasting since they can capture complex nonlinear patterns and long-term dependencies [5, 6]. While CNNs were initially developed for image processing tasks, they can be repurposed to capture local structures and temporal dynamics within time series data [7]. LSTMs, a variant of Recurrent Neural Networks, are particularly well-suited for capturing long-range dependencies in financial sequential data [8].

In addition, ensemble learning methods such as Random Forest are also able to improve model robustness. This is due to their interpretability, resistance to overfitting, and ability to rank the importance of features [9].

Recently, hybrid models that combine statistical models with deep learning models have become increasingly prominent in the field of financial prediction. These models aim to leverage the interpretability and stability of statistical methods alongside the capability of deep learning models to

capture nonlinear relationships and hidden patterns. For example, ARMA models can effectively describe volatility clustering, while LSTMs and CNNs can detect subtle price movements that linear or variance-based models may miss. By averaging or weighting the outputs of these different methods, hybrid frameworks have shown better forecasting performance [10]. The study of He et al. reveals that the hybrid ARMA-CNNLSTM model outperforms the traditional ARMA model, achieving reductions in the MAPE of test results by 3.15%, 17.13%, and 1.55% across three distinct test datasets [10].

The purpose of this research is to assess and contrast the performance of different modeling approaches in predicting the S&P 500, FTSE 100, and Nikkei 225 stock indices. The models used include the traditional AR-GARCH model, machine learning models (CNN, LSTM, and Random Forest), as well as a hybrid model that integrates predictions from all these approaches.

2. Methods

2.1. Dataset Description

This study selects three major stock indices from different economies to evaluate global market behavior. Representing the UK economy, the FTSE 100 Index ranks the top 100 companies on the London Stock Exchange by their market value and functions as a major indicator reflecting the overall health of Britain's stock market. The S&P 500 Index (United States) includes 500 leading companies from major sectors of the US economy, representing about 75% of the total market capitalization of US equities. The Nikkei 225 Index (Japan) is a price-weighted index comprising 225 prominent forms traded on the Tokyo Stock Exchange and plays the role of measuring of the Japanese stock market. Together, these indices represent central components of the UK, US, and Japanese economies, providing a diverse and independent dataset to test model generalization across different market environments.

The data used in this study were obtained from Yahoo Finance and include weekly closing prices for each index from January 2000 to January 2024. To ensure consistency in the time series and address missing values, linear interpolation was applied to fill gaps in the weekly data. The price series was then transformed into a stationary log-return series by calculating log returns. This transformation stabilizes variance and reduces level trends in the series, which is a common requirement for applying AR and GARCH models [11]. Based on the return series, additional predictive features were constructed, including lagged returns as well as technical indicators such as moving averages and recent volatility. These features enable the model to identify time-based relationships and recurring trends within the dataset.

The full dataset spans 24 years, with the period from 2000 to 2020 used for training and 2021 to 2024 used for testing. This time-based split (approximately 80% training set and 20% test set) ensures that the model is evaluated on data not used during estimation, and simulates a realistic forecasting scenario. By applying the same procedure independently to the S&P 500, FTSE 100, and Nikkei 225, three parallel datasets were obtained. This allows for assessing the robustness of different modeling methods. If a model performs well across all three markets, it is more likely to generalize effectively in diverse economic conditions.

2.2. Modelling

2.2.1. AR-GARCH

The AR-GARCH model holds significant theoretical and practical value in the analysis of financial time series. It effectively captures two key components commonly observed in asset returns, which are the linear autoregressive mean and the generalized autoregressive conditional heteroskedasticity variance.

The AR component models linear autocorrelation in returns, while the GARCH component captures time-varying volatility. That is, large past shocks typically lead to higher current conditional variance a phenomenon known as volatility clustering [12]. From an economic perspective, this reflects the empirical observation that periods of high volatility tend to persist.

Assuming normally distributed errors, the AR-GARCH model is statistically efficient and provides an explicit expression for forecasting conditional variance. Nevertheless, its performance depends on correct model specification. If returns deviate significantly from normality or exhibit nonlinear patterns beyond volatility clustering, the AR-GARCH model may produce biased or inefficient forecasts [13]. Despite these limitations, AR-GARCH remains a valuable benchmark for capturing both mean and variance dynamics in financial return series, offering insights into memory effects in both dimensions [14].

2.2.2. CNN

In this research, a CNN in one dimension is utilized to capture localized temporal features from the return series. A major advantage of CNNs lies in their ability to combine convolution and pooling layers to gradually expand the receptive field while reducing dimensionality. This enables the model to flexibly identify market dynamics over varying time horizons, ranging from brief quick-term movements to longer-term trends [15].

CNNs are characterized by two main inducive biases, which are locality and translation equivariance [16]. By applying the same filter across different time steps, the network can consistently detect specific patterns, regardless of their position in the sequence. This property is particularly useful in financial data analysis, where similar price patterns often recur over time.

2.2.3. LSTM

As a specialized form of RNN, LSTM is able to incorporate memory cells and gating mechanisms to efficiently recognize long-range patterns in sequential data while mitigating vanishing and exploding gradient issues [17]. These gating mechanisms regulate the information flow in the cell state, enabling precise management of memory in the long run [17].

LSTM, a specialized type of RNN, incorporates memory cells along with input, output, and forget gates to regulate information flow within the cell state. This architecture enables the network to efficiently recognize long-range patterns in sequential data while mitigating vanishing and exploding gradient issues, thereby ensuring more stable and accurate modeling in the long run [17].

In forecasting financial time series, LSTM networks excel at identifying temporal autocorrelations and state transition patterns within return sequences. Through its unique gating mechanism, the model is theoretically capable of learning complex temporal patterns that include quick shifts and long-duration tendencies. However, practically, LSTM networks often require large amounts of training data, and they are prone to overfitting when data is limited. To address this, this study applies dropout regularization and hyperparameter tuning to improve its general performance [18].

2.2.4. Random Forest

As an ensemble technique, random forest improves predictive results by integrating the outputs of numerous decision trees. It relies on two key techniques: bootstrap sampling and random feature selection. Each base learner is fitted using a bootstrapped sample of the original dataset, and at each decision node, only a random portion of the available features is evaluated. Eventually, their outputs are averaged to form the prediction.

In financial time series forecasting, Random Forest offers advantages such as capturing nonlinear relationships and handling various features without requiring strict assumptions [19]. However, it lacks explicit modeling of temporal dynamics and treats time as a regular feature, limiting its ability to represent market volatility [20]. Despite this, the robustness and ability of Random Forest to detect complex patterns still make it a useful and complementary tool in empirical studies.

2.3. Hybrid Strategy

After fitting four individual models, their corresponding forecasts are combined into an equally weighted ensemble. This hybrid approach is employed based on two key considerations that need to be demonstrated. By partially offsetting random errors from different models, averaging may reduce the overall variance of prediction errors [21]. In cases where model errors are uncorrelated or weakly correlated, the hybrid variance could be significantly lower than that of any individual model [22]. Averaging may also mitigate model specification bias. If one model performs poorly in a specific context, others could compensate to enhance the overall performance.

3. Results and Discussion

3.1. Results

After modelling, the test set is used to generate five separate predictions in terms of log-returns. These predictions are then converted into closing prices, enabling the differences between models to become more apparent. To identify the best-performing model, the root mean squared error (MSE), mean absolute error (MAE), and mean absolute percentage error (MAPE) are computed and summarized in Tables 1 to 3.

3.2. Discussion

Table 1. Model Performance for FTSE 100 Dataset

	RMSE	MAE	MAPE
AR-GARCH	231.15	174.54	2.45%
LSTM	250.02	156.77	2.34%
CNN	253.02	197.20	2.78%
Random Forest	290.22	201.69	2.71%
Hybrid	227.17	166.87	2.30%

The empirical results indicate that different models show varying predictive performance across the indices. For the FTSE 100 index, the hybrid model performs best, achieving the lowest RMSE (227.17) and MAPE (2.30%), slightly outperforming the AR-GARCH model (RMSE = 231.15, MAPE = 2.45%). Despite having the lowest MAE value at 156.77, the LSTM model yields a higher RMSE compared to the AR-GARCH model. These results suggest that the LSTM architecture, despite its stability, is prone to producing large prediction errors at times. This finding aligns with the LSTM's ability to capture nonlinear dynamics and its tendency to overfit during periods of high volatility.

Table 2. Model Performance for S&P 500 Dataset

	RMSE	MAE	MAPE
AR-GARCH	211.32	145.06	3.74%
LSTM	174.89	125.86	3.19%
CNN	229.84	154.53	4.02%
Random Forest	204.20	169.18	4.06%
Hybrid	181.79	130.75	3.32%

For the S&P 500 index, the LSTM model shows clear advantages, with the lowest RMSE (174.89), MAE (125.86), and MAPE (3.19%). This highlights the ability of the model to detect and represent the temporal complexities of the US market, which is characterized by high liquidity and rapid information absorption. The hybrid model ranks second (RMSE = 181.79, MAE = 130.75, MAPE = 3.32%), reflecting its strength in mitigating the influence of outliers through the aggregation of multiple model predictions.

Table 3. Model Performance for Nikkei 225 Dataset

	RMSE	MAE	MAPE
AR-GARCH	1367.72	1085.90	3.04%
LSTM	1455.61	1128.40	3.17%
CNN	1503.75	1127.52	3.17%
Random Forest	1627.05	1251.28	3.54%
Hybrid	1394.01	1061.78	2.98%

A similar pattern of the FTSE 100 index could also be found in the Nikkei 225 index. The hybrid model has an outstanding performance in terms of MAE and MAPE (MAE = 1061.78, MAPE = 2.98%). However, the AR-GARCH model has a lower RMSE score (1367.72). This is reasonable, as the Japanese market, which is affected by long-term deflationary pressure and policy interventions, may exhibit stronger linear dependence and more persistent volatility.

While CNNs showed promise in early financial applications, our results indicate limitations in standalone forecasting. The relatively high errors (RMSE = 253.02, 229.84, or 1503.75) suggest CNNs may require additional feature engineering or hybrid architectures for financial time series. Recent work by Zeng et al. confirms that pure CNNs often underperform compared to sequential models like LSTMs for price prediction [15].

Random Forest demonstrated the highest errors among all models (MAE = 201.69, 169.18, or 1251.28), consistent with findings by Masini et al. that tree-based methods struggle with temporal dependencies [19].

The CNN and Random Forest models perform relatively poorly across all three scenarios, possibly as a result of insufficient capabilities to represent the sequential dependencies and volatility clustering characteristic of stock index movements. Nevertheless, they still add value to the hybrid model, as their inclusion helps reduce overall prediction variance through averaging. Moreover, these models may detect some complex and non-sequential short-term patterns that other models omit.

3.3. Limitations and Future Work

In real-world applications, stock index volatility often responds asymmetrically to positive and negative changes, limiting the effectiveness of the AR-GARCH model. To better address this leverage effect, a Threshold GARCH (TGARCH) model could be employed in place of the standard GARCH model. TGARCH models outperform traditional GARCH models in capturing these asymmetries, as they explicitly model the differential effects of upward and downward shocks on volatility, offering a more accurate representation of financial market behavior [23]. Empirical evidence also shows that TGARCH better captures leverage effects and asymmetric information flow in trading volume dynamics, which standard GARCH models tend to miss [24]. Moreover, the hybrid model currently uses an equal-weighting approach, assuming all models contribute equally, which may not hold in practice. Future research may explore adaptive weighting strategies to enhance overall predictive performance.

4. Conclusion

This research aims to compare the performance of AR-GARCH, CNN, LSTM, Random Forest, and their hybrid combinations in forecasting weekly stock indices (S&P 500, FTSE 100, and Nikkei 225) from 2000 to 2024. Evaluation using RMSE, MAE, and MAPE revealed that each model performed differently depending on the market characteristics.

The AR-GARCH model showed strong performance in capturing linear dynamics and volatility, particularly in the Nikkei 225. LSTM outperformed other models for the S&P 500, illustrating its effectiveness in learning long-range time dependencies. CNN or Random Forest performed less consistently, with CNN suffering from weak temporal modeling and Random Forest limited by its

inability to capture sequential patterns. The hybrid model, which averaged predictions from all four models, consistently performed best overall. It achieved the lowest errors in both the FTSE 100 and Nikkei 225, and ranked second for the S&P 500, confirming its robustness across diverse markets.

Overall, combining models helps mitigate individual weaknesses and enhances forecast accuracy. Future work should explore adaptive ensemble weighting and incorporate models that capture asymmetric volatility, such as TGARCH, to further improve forecasting reliability in real-world applications.

References

[1] R.F. Engle, Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation, Econometrica 50 (1982) 987–1007.

[2] T. Bollerslev, Generalized autoregressive conditional heteroskedasticity, J. Econometrics 31 (1986) 307–327.

[3] P.H. Franses, D. Van Dijk, Forecasting stock market volatility using (non-linear) GARCH models, J. Forecast. 15 (1996) 229–235.

[4] S.J. Taylor, Asset Price Dynamics, Volatility, and Prediction, Princeton University Press, Princeton, 2011.

[5] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput. 9 (1997) 1735–1780.

[6] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (2015) 436–444.

[7] L. Zhang, C. Aggarwal, G.J. Qi, Stock price prediction via discovering multi-frequency trading patterns, in: Proc. 23rd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min., 2017, pp. 2141–2149.

[8] W. Bao, J. Yue, Y. Rao, A deep learning framework for financial time series using stacked autoencoders and long-short term memory, PLoS One 12 (2017) e0180944.

[9] S. Gu, B. Kelly, D. Xiu, Empirical asset pricing via machine learning, Rev. Financ. Stud. 33 (2020) 2223–2273.

[10] K. He, Q. Yang, L. Ji, J. Pan, Y. Zou, Financial time series forecasting with the deep learning ensemble model, Mathematics 11 (2023) 1054.

[11] J. Janczura, A. Puć, ARX-GARCH probabilistic price forecasts for diversification of trade in electricity markets—variance stabilizing transformation and financial risk-minimizing portfolio allocation, Energies 16 (2023) 807.

[12] M.L. Bianchi, G. De Luca, G. Rivieccio, CoVaR with volatility clustering, heavy tails and non-linear dependence, arXiv preprint arXiv:2009.10764 (2020).

[13] J. Michańków, Ł. Kwiatkowski, J. Morajda, Combining deep learning and GARCH models for financial volatility and risk forecasting, arXiv preprint arXiv:2310.01063 (2023).

[14] J. Wei, S. Yang, Z. Cui, Integrated GARCH-GRU in financial volatility forecasting, arXiv preprint arXiv:2504.09380 (2025).

[15] Z. Zeng, R. Kaur, S. Siddagangappa, S. Rahimi, T. Balch, M. Veloso, Financial time series forecasting using CNN and transformer, arXiv preprint arXiv:2304.04912 (2023).

[16] Z. Wang, L. Wu, Theoretical analysis of the inductive biases in deep convolutional networks, Adv. Neural Inf. Process. Syst. 36 (2023) 74289–74338.

[17] K. Greff, R.K. Srivastava, J. Koutník, B.R. Steunebrink, J. Schmidhuber, LSTM: A search space odyssey, IEEE Trans. Neural Netw. Learn. Syst. 28 (2016) 2222–2232.

[18] O.B. Sezer, M.U. Gudelek, A.M. Ozbayoglu, Financial time series forecasting with deep learning: a systematic literature review: 2005–2019, Appl. Soft Comput. 90 (2020) 106181.

[19] R.P. Masini, M.C. Medeiros, E.F. Mendes, Machine learning advances for time series forecasting, J. Econ. Surv. 37 (2023) 76–111.

[20] M. Zhao, X. Zhang, J. Appiah, M.D. Fontaine, Travel time reliability prediction using random forests, Transp. Res. Rec. 2678 (2024) 531–545.

[21] K. Kakade, I. Jain, A.K. Mishra, Value-at-Risk forecasting: a hybrid ensemble learning GARCH-LSTM based approach, Resour. Policy 78 (2022) 102903.

[22] A. Mahmoud, A. Mohammed, Leveraging hybrid deep learning models for enhanced multivariate time series forecasting, Neural Process. Lett. 56 (2024) 223.

[23] M.H. Ibrahim, A.Z. Baharumshah, M.S. Habibullah, Comparing the performances of GARCH-type models in capturing the stock market volatility in Malaysia, J. Appl. Sci. 14 (2014) 3557–3564.

[24] W.K. Wong, J. Li, Modeling and forecasting trading volume index: GARCH versus TGARCH approach, Q. Rev. Econ. Finance 50 (2010) 319–329.