Jurnal Statistika Dan Komputasi (STATKOM), Vol. 5 No. 1 (June 2026)
ISSN : 2963-038X (print), ISSN : 2963-0398
DOI : https://doi.org/10.32665/statkom.v5i1.6463

STATKOM
JURNAL STATISTIKA DAN KOMPUTASI

Forecasting the Market Dynamics of the Invesco QQQ Trust Using Interpretable Econometric and Deep Learning Approaches

Aris Rakhmadi[1], Rakha Muhammad Fatwa[2], Ahmad Mardalis[3]

[1,2]Department of Informatics Engineering, Universitas Muhammadiyah Surakarta
[3]Department of Management, Universitas Muhammadiyah Surakarta
Email: aris.rakhmadi@ums.ac.id

Submitted: April 20, 2026 ***Revised:*** June 18, 2026 ***Accepted:*** June 23, 2026

Abstract

Background: *Financial time-series forecasting requires models that balance predictive accuracy and interpretability, while the advantage of deep learning over classical econometric approaches remains uncertain for highly liquid financial instruments.*
Objective: *This study compares interpretable econometric and deep learning approaches to identify the most suitable model for forecasting the closing price and trading volume of the Invesco QQQ Trust.*
Methods: *The study used 1,258 daily observations from 2019 to 2024 obtained from Yahoo Finance. Closing price and trading volume were treated as separate forecasting targets. Separate ARIMA(7,1,0) models were applied to each target and compared with Vanilla LSTM, Bidirectional LSTM, and Stacked LSTM. Performance was evaluated using RMSE, MAE, and MAPE.*
Results: *ARIMA recorded the lowest errors for both targets, with MAPE values of 0.82% for closing price and 19.96% for trading volume. Vanilla LSTM was the best-performing deep learning architecture, while Bidirectional and Stacked LSTM produced larger errors. The higher volume error indicates that trading activity was more difficult to forecast than closing-price movements.*
Conclusion: *Within the examined QQQ data, period, and model configurations, ARIMA provided more accurate forecasts than the evaluated LSTM architectures. Greater model complexity did not necessarily improve predictive performance.*

Keywords: *Time-Series Forecasting, Invesco QQQ Trust, ARIMA, LSTM, Financial Forecasting.*

Abstrak

Latar Belakang: Peramalan deret waktu keuangan membutuhkan model yang menyeimbangkan akurasi prediksi dan interpretabilitas, sedangkan keunggulan deep learning dibandingkan pendekatan ekonometrik klasik masih belum pasti pada instrumen keuangan yang sangat likuid.
Tujuan: Penelitian ini membandingkan pendekatan ekonometrik yang dapat diinterpretasikan dan deep learning untuk mengidentifikasi model yang paling sesuai dalam meramalkan harga penutupan dan volume perdagangan Invesco QQQ Trust.
Metode: Penelitian menggunakan 1.258 observasi harian periode 2019–2024 yang diperoleh dari Yahoo Finance. Harga penutupan dan volume perdagangan diperlakukan sebagai target peramalan terpisah. Model ARIMA(7,1,0) diterapkan secara terpisah pada setiap target dan dibandingkan dengan Vanilla LSTM, Bidirectional LSTM, dan Stacked LSTM. Kinerja dievaluasi menggunakan RMSE, MAE, dan MAPE.
Hasil: ARIMA menghasilkan kesalahan terendah pada kedua target, dengan MAPE sebesar 0,82% untuk harga penutupan dan 19,96% untuk volume perdagangan. Vanilla LSTM merupakan arsitektur deep learning terbaik, sedangkan Bidirectional dan Stacked LSTM menghasilkan kesalahan lebih besar. Kesalahan volume yang lebih tinggi menunjukkan bahwa aktivitas perdagangan lebih sulit diramalkan dibandingkan pergerakan harga penutupa.
Kesimpulan: Pada data QQQ, periode, dan konfigurasi model yang diuji, ARIMA menghasilkan peramalan lebih akurat daripada arsitektur LSTM. Kompleksitas model yang lebih tinggi tidak selalu meningkatkan kinerja prediktif.

Kata kunci: Peramalan Deret Waktu, Invesco QQQ Trust, ARIMA, LSTM, Peramalan Keuangan.

INTRODUCTION

Financial markets function as fundamental mechanisms for capital allocation, risk sharing, and price discovery in modern economies. Through continuous trading activities, asset prices aggregate dispersed information held by heterogeneous investors and reflect collective expectations about future economic conditions (Apriantoro et al., 2024). The efficiency with which markets incorporate new information has long been a central question in financial economics. Within this context, Exchange-Traded Funds (ETFs) have emerged as important financial instruments that combine portfolio diversification, high liquidity, and real-time price discovery. Their rapid expansion has positioned ETFs as key vehicles for gaining exposure to diversified market indices while maintaining the flexibility of intraday trading (Divelgama et al., 2026).

The growing dominance of ETFs in global capital markets raises important questions about the evolution of their price and trading-volume dynamics. Because ETFs trade continuously on exchanges, their market behavior provides a rich environment for studying short-term predictability and information diffusion (Álvarez et al., 2026; Novotny & Hajek, 2026). Price movements and trading volumes often reflect shifts in investor sentiment, macroeconomic signals, and broader perceptions of market risk. In highly liquid markets, financial theory suggests that new information should be rapidly incorporated into prices, limiting the scope for predictable patterns. However, empirical evidence often reveals short-term dependencies in financial time series, suggesting that certain market dynamics may remain predictable.

Understanding whether such dependencies can be systematically forecasted is both theoretically and practically relevant. From a theoretical perspective, predictable patterns may indicate temporary deviations from informational efficiency (Naifar, 2026). From a practical standpoint, improved forecasting models can support portfolio allocation, trading strategies, and risk management decisions. In particular, the joint modeling of asset prices and trading volumes provides deeper insight into market behavior, as trading volume is often interpreted as a proxy for information arrival and trading intensity. Despite the importance of these interactions, the literature has not reached a consensus regarding the most appropriate modeling framework for capturing the intertwined dynamics of prices and volumes in financial markets.

Recent advances in machine learning and deep learning have introduced flexible modeling frameworks that can capture nonlinear temporal dependencies in financial time-series data. Recurrent neural network architectures, particularly Long Short-Term Memory (LSTM) networks, have gained attention for their ability to model complex sequential relationships and long-range dependencies (Z. Chen et al., 2026; Zeng et al., 2026). Numerous empirical studies report improvements in predictive accuracy when applying such techniques to financial forecasting tasks (Y. Chen, 2025; Lee et al., 2026). However, improvements in statistical forecasting accuracy do not necessarily translate into economically meaningful insights, especially when model complexity reduces interpretability and increases model risk.

A critical gap in the literature concerns the economic interpretation of model complexity. While many studies compare forecasting accuracy across different algorithms, relatively few investigate whether additional computational complexity improves our understanding of financial market behavior (Frimpong & Mamuti, 2026). This issue is particularly relevant in highly liquid ETF markets, where rapid information assimilation may limit the benefits of nonlinear modeling approaches. Complex models may yield marginal gains in predictive accuracy, but these gains may come at the cost of transparency, robustness, and interpretability.

Another limitation in existing research lies in the scope of model comparison. Many empirical studies contrast a single econometric model with a single machine learning architecture, often without systematically evaluating alternative neural network structures. Moreover, much of the literature focuses primarily on univariate price prediction, neglecting the joint behavior of prices and trading volumes (Maryam et al., 2022). This omission restricts the ability to examine broader market dynamics and liquidity-related mechanisms that influence financial market activity.

ETFs have become increasingly important financial instruments because they combine portfolio diversification, intraday trading flexibility, liquidity, and transparency. By replicating broad indices or sector-specific portfolios, ETFs provide efficient market exposure while also serving as useful instruments for examining price discovery, market efficiency, and liquidity formation (Utami & Irawati, 2021). The dynamics of ETF prices and trading volumes reflect the interaction among investor behavior, macroeconomic signals, and broader market sentiment. Since ETFs are continuously traded during market hours, their price and volume movements can represent real-time information diffusion in financial markets (Geng et al., 2025; Ozcelebi et al., 2025).

Traditional econometric models remain widely used in financial time-series forecasting because of their transparency, interpretability, and strong statistical foundation. The Autoregressive Integrated Moving Average (ARIMA) model is one of the most common approaches for forecasting economic and financial variables because it can capture linear temporal dependencies and short-term autocorrelation structures (Sahu et al., 2026). In empirical finance, ARIMA has been applied to various financial indicators such as stock prices, exchange rates, commodity prices, and other time-series variables characterized by structured temporal movements. Its interpretability makes it particularly useful when forecasting results must be explained and evaluated in relation to market behavior.

Nevertheless, linear econometric models may have limitations when financial data contain nonlinear patterns, structural breaks, or complex temporal interactions. These limitations have encouraged the use of machine learning and deep learning approaches in financial forecasting. Among deep learning models, Long Short-Term Memory (LSTM) networks have gained considerable attention because their gated memory mechanisms allow them to retain relevant historical information and capture long-range temporal dependencies. Empirical studies have reported that LSTM-based models can outperform traditional statistical approaches in modeling nonlinear dependencies and sequential patterns in financial time-series data (Silfiani et al., 2024).

Several LSTM variants have been proposed to improve forecasting performance in sequential data. Vanilla LSTM is commonly used as a baseline architecture, while Bidirectional LSTM processes temporal sequences in both forward and backward directions to improve contextual learning. Stacked LSTM uses multiple LSTM layers to learn deeper hierarchical temporal representations within financial data. These architectures have been applied in various financial forecasting tasks, including stock price prediction, volatility modeling, and asset return forecasting (Li, 2025; Su & Wang, 2026). However, greater architectural complexity does not always guarantee better predictive performance, particularly when the available dataset is limited or when the market exhibits relatively structured temporal behavior.

Despite the growing literature on financial forecasting, several gaps remain. First, many previous studies focus primarily on individual stock prediction. In contrast, relatively fewer studies investigate ETFs as distinct financial instruments with unique characteristics such as high liquidity, index replication, and continuous intraday trading (Bashir et al., 2026; Xu et al., 2026). Second, many studies compare only one econometric model with one machine learning model, which limits the understanding of how model complexity affects forecasting performance (Mo, 2025; Rios-Vazquez & Portela-Maseda, 2026). Third, much of the existing literature emphasizes univariate price forecasting and gives limited attention to the joint dynamics between price and trading volume (Ajay et al., 2026; Solanki & Jha, 2026; Tsoku & Makatjane, 2026).

Against this background, this study investigates whether increasing model complexity leads to meaningful improvements in forecasting ETF market dynamics (Amankwah et al., 2026; Chichernea et al., 2026). Specifically, the research compares an interpretable econometric benchmark with several deep learning architectures in forecasting two ETF market indicators: closing price and trading volume. The empirical analysis focuses on the Invesco QQQ Trust (QQQ), one of the most actively traded ETFs tracking the Nasdaq-100 index. Due to its high liquidity and continuous trading activity, the QQQ ETF provides an appropriate setting for examining short-term predictability and information diffusion in a mature financial market.

The novelty of this research lies not only in comparing econometric and deep learning approaches but also in interpreting forecasting performance within the broader framework of financial market efficiency and model interpretability. Rather than treating predictive accuracy as the sole evaluation criterion, this study assesses the relationships among model complexity, market structure, liquidity, and informational efficiency (Nuryakin et al., 2021). By highlighting the trade-off between interpretability and predictive flexibility, the analysis provides insights relevant to investors, analysts, and policymakers concerned with model risk and practical applicability. Consequently, the findings contribute to the ongoing debate regarding the appropriate balance between traditional econometric rigor and emerging computational approaches in financial market analysis.

METHOD

Research Design

This study employed a quantitative comparative forecasting design using secondary financial time-series data. The research was designed to evaluate and compare the forecasting performance of interpretable econometric and deep learning approaches in modeling ETF market dynamics. The econometric approach was represented by the Autoregressive Integrated Moving Average (ARIMA) model. In contrast, the deep learning approaches consisted of Vanilla Long Short-Term Memory (LSTM), Bidirectional LSTM, and Stacked LSTM.

The study focused on forecasting two market variables of the Invesco QQQ Trust: closing price and trading volume. These variables were treated as separate forecasting targets representing market valuation and trading intensity, respectively. For the econometric benchmark, a separate ARIMA model was estimated for each target series. The use of a comparative forecasting design allowed the study to examine whether increasing model complexity improves prediction accuracy in a highly liquid ETF market.

The statistical evaluation was conducted using an out-of-sample forecasting framework. All models were trained and tested using the same dataset, target variables, data partitioning scheme, and evaluation metrics to ensure a fair comparison. Forecasting performance was measured using Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). Lower values of these metrics indicate better forecasting performance. This design was used to assess not only predictive accuracy but also the trade-off between model interpretability and model complexity.

Population and Sample

The population of this study refers to daily market data of highly liquid ETFs, particularly ETFs that represent active trading behavior, market liquidity, price discovery, and continuous information diffusion in financial markets. This population was selected because the study focuses on forecasting ETF market dynamics through the interaction between price movement and trading activity.

The sample used in this study was the Invesco QQQ Trust (QQQ), an ETF that tracks the Nasdaq-100 index and is among the most actively traded ETFs globally. QQQ was selected because it represents a highly liquid ETF market and provides a suitable empirical context for examining short-term forecasting behavior in a mature financial market.

The dataset consisted of 1,258 daily QQQ market observations from 2019 to 2024 obtained from Yahoo Finance and arranged chronologically. The forecasting analysis examined two target series: daily closing price and trading volume. The closing price represented ETF market valuation, whereas trading volume represented trading intensity and liquidity-related market activity. Forecasting performance was calculated and reported separately for the two targets.

Sampling Techniques

This study used purposive sampling to select the research object. The Invesco QQQ Trust (QQQ) was selected because it met several criteria relevant to the research objective. First, QQQ is a highly liquid ETF that represents active trading behavior in a mature financial market. Second, it tracks the Nasdaq-100 index, making it suitable for observing the dynamics of technology-oriented market movements. Third, QQQ provides complete historical daily data for the 2019–2024 observation period, including closing price and trading volume, which are the main variables in this study.

The unit of analysis in this study was the daily trading observation. After the QQQ ETF was selected, all available daily observations within the defined period were included in the dataset. Therefore, the study did not use random sampling because the data were not collected from individual respondents but from archival financial records arranged chronologically. This approach was appropriate for time-series forecasting because maintaining the natural temporal order of the data is essential to avoid look-ahead bias.

The selected sample consisted of 1,258 daily observations obtained from Yahoo Finance. These observations were organized chronologically and used to forecast the closing-price and trading-volume series separately. The use of purposive sampling ensured that the selected ETF was relevant to the study's focus on forecasting market dynamics in a highly liquid ETF environment.

Research Subjects

The research subject in this study was secondary archival financial data of the Invesco QQQ Trust (QQQ) obtained from the Yahoo Finance database. The data were collected at a daily frequency for the period from 2019 to 2024 and arranged chronologically to preserve their temporal ordering. The dataset consisted of daily market variables, including date, open price, high price, low price, closing price, adjusted closing price, and trading volume. A sample structure of the ETF market dataset is presented in Table 1.

Table 1. Sample Structure of the ETF Market Dataset

Date	Low	High	Open	Close	Adj Close	Volume
2024-02-15	431.33	434.98	433.92	434.51	434.51	38796100
2024-02-16	429.85	434.99	434.89	430.57	430.57	53661500
2024-02-20	423.5	430.08	428.55	427.32	427.32	53999500
2024-02-21	421.63	425.7	424.55	425.61	425.61	50179700
2024-02-22	433.71	439.12	434.49	438.07	438.07	53751500

Table 1 shows the structure of the market data used in this study. Although the dataset contains several daily market attributes, the main variables analyzed in the forecasting process were closing price and trading volume. The closing price was used to represent ETF market valuation, while trading volume was used to represent trading intensity and liquidity-related market information.

To provide an overview of the research subject, Figure 1 presents the historical closing prices of the Invesco QQQ ETF during the observation period. The figure illustrates the general price movement from 2019 to 2024, including long-term trends and periods of market fluctuation.

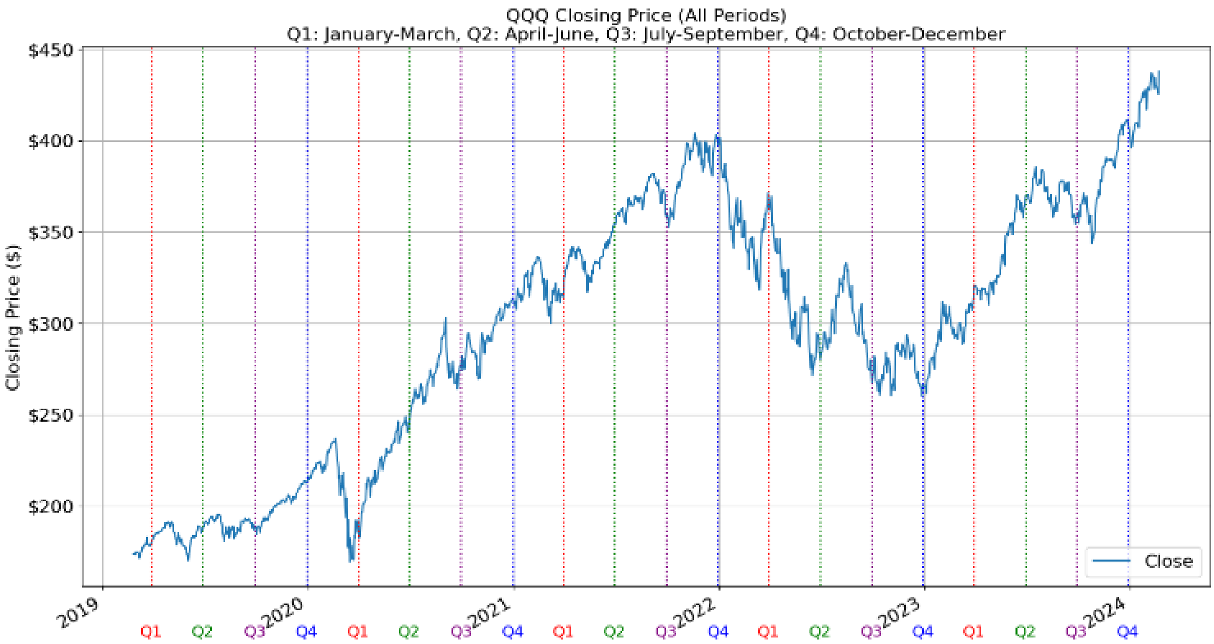

Figure 1. Historical Closing Prices of the Invesco QQQ ETF (2019–2024)

Figure 2 presents the historical trading volume of the Invesco QQQ ETF during the same observation period. The volume series shows higher variability than the closing price series, with frequent spikes that may reflect changes in trading intensity, investor participation, and information arrival in the market.

In this study, closing price and trading volume were retained as two complementary market indicators. The two series were forecast and evaluated separately, enabling the analysis to compare model performance for market valuation and trading activity without implying that the ARIMA model estimated their joint dependence.

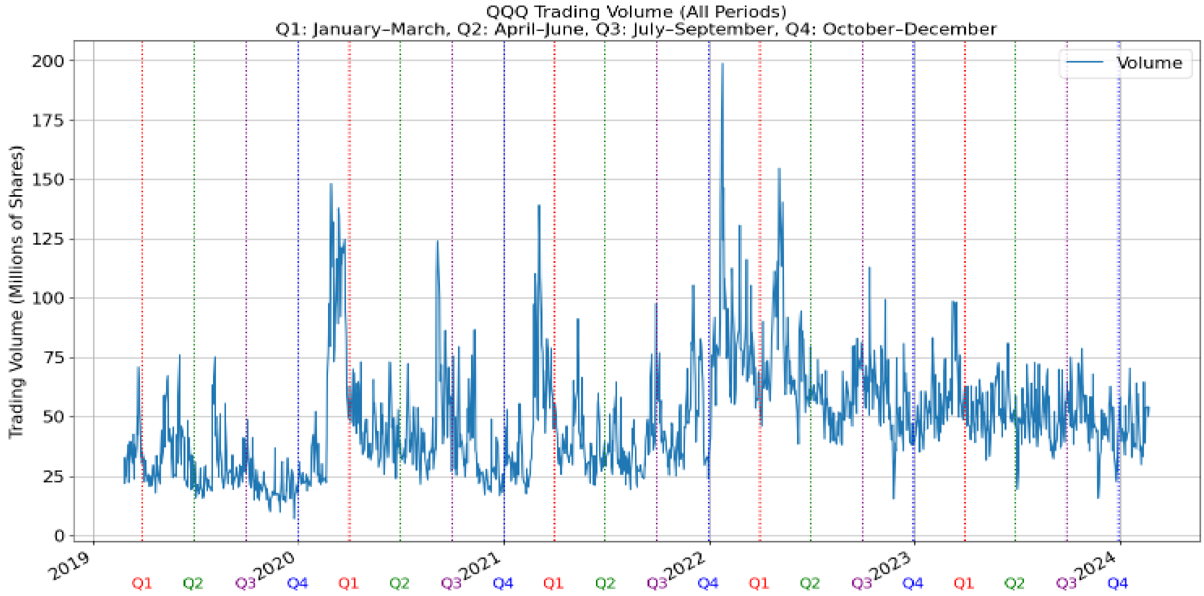

Figure 2. Historical Trading Volume of the Invesco QQQ ETF (2019–2024)

Measurement Instrument Design

This study did not use questionnaires, interviews, or direct observation instruments because the data were obtained from secondary archival sources. The measurement instrument used in this study was an archival data collection procedure based on the Yahoo Finance database. Yahoo Finance was used to obtain daily historical market data of the Invesco QQQ Trust (QQQ) for the 2019–2024 observation period.

The archival data collection process consisted of several steps. First, the QQQ ticker was selected as the research object. Second, the observation period was defined from 2019 to 2024. Third, daily historical market data were collected, including date, open price, high price, low price, closing price, adjusted closing price, and trading volume. These variables correspond to the dataset structure previously shown in Table 1. Fourth, the collected data were checked, arranged chronologically, and prepared as a multivariate time-series dataset for forecasting analysis.

The main measured variables were closing price and trading volume. The closing price was used to measure ETF market valuation, while trading volume was used to measure trading intensity and liquidity-related market activity. Because the data were based on recorded financial transactions, the measurement process relied on archival market records rather than subjective responses. This approach ensured that the analysis used objective numerical data suitable for time-series forecasting.

Data Analysis Techniques

The data analysis was conducted through a structured time-series forecasting procedure. First, daily market data of the Invesco QQQ Trust (QQQ) were collected from Yahoo Finance and arranged chronologically. Second, closing price and trading volume were selected as the two forecasting targets. The closing price represented ETF market valuation, whereas trading volume represented trading intensity and liquidity-related market information. Each target series was forecast and evaluated separately to ensure that performance differences were directly comparable across models.

Before model development, the dataset was divided sequentially into training, validation, and testing subsets. The training set consisted of 880 observations, the validation set consisted of 188 observations, and the testing set consisted of 190 observations. The chronological splitting technique was used to ensure that the models were trained using past observations and evaluated using unseen future data. This procedure was applied to reduce potential look-ahead bias in the forecasting process.

Four forecasting approaches were evaluated in this study. For the econometric benchmark, separate ARIMA models were applied to the closing-price and trading-volume series because standard ARIMA is a univariate forecasting model. Both target series used the ARIMA(7,1,0) specification, consisting of an autoregressive order of seven, first-order differencing, and no moving-average component. Accordingly, the forecasting results for closing price and trading volume were calculated and reported separately.

The deep learning approaches consisted of Vanilla LSTM, Bidirectional LSTM, and Stacked LSTM. Vanilla LSTM and Bidirectional LSTM used two recurrent layers with 64 hidden units, whereas Stacked LSTM used three recurrent layers with 64 hidden units per layer. The LSTM models were trained using the Adam optimizer with a learning rate of 0.001, mean squared error as the loss function, a mini-batch size of 32, and 150 training epochs. All models were implemented in Python using TensorFlow and evaluated using the same chronological data partitions and forecasting metrics. The complete model configurations are presented in Table 2.

Table 2. Model Parameter Settings

Model	Configuration Details
ARIMA	Autoregressive lag (p): 7
	Degree of differencing (d): 1
	Moving average component (q): 0
Vanilla LSTM	Number of LSTM layers: 2
	Hidden neurons per layer: 64
	Mini-batch size: 32
	Training epochs: 150
Bidirectional LSTM	Number of bidirectional layers: 2
	Hidden neurons: 64
	Mini-batch size: 32
	Training epochs: 150
Stacked LSTM	Total stacked layers: 3
	Hidden neurons per layer: 64
	Mini-batch size: 32
	Training epochs: 150

The research workflow is presented in Figure 3. The analysis began with the collection of daily historical market data of the Invesco QQQ Trust from Yahoo Finance. The dataset was then organized chronologically and prepared as a multivariate time-series dataset consisting of closing price and trading volume. After preprocessing, the data were divided sequentially into training, validation, and testing subsets to preserve the temporal structure of the observations and reduce potential look-ahead bias.

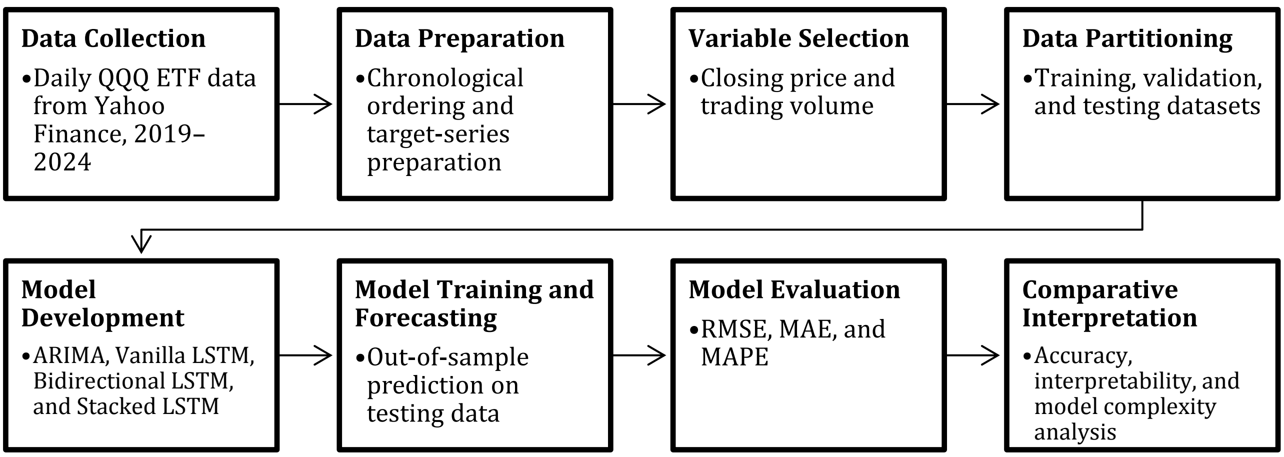

Figure 3. Methodological Framework for Comparative ETF Market Forecasting

The forecasting stage involved the development of one econometric model and three deep learning models. The ARIMA model was used as the interpretable econometric benchmark, while Vanilla LSTM, Bidirectional LSTM, and Stacked LSTM were used to represent deep learning approaches. Each model was trained using the same data partition and evaluated on the testing dataset. The final stage involved comparing model performance using RMSE, MAE, and MAPE, followed by an interpretation of the forecasting results in relation to model accuracy, interpretability, and complexity.

Forecasting performance was evaluated using three statistical error metrics: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). RMSE was used to measure the magnitude of forecasting error with greater sensitivity to large errors. MAE was used to calculate the average absolute difference between actual and predicted values. MAPE was used to express the forecasting error in percentage form, making it easier to interpret the relative prediction error.

The formulas used in this study are as follows

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(Y_t - \hat{Y}_t)^2}$$

$$MAE = \frac{1}{n}\sum_{t=1}^{n}\left|Y_t - \hat{Y}_t\right|$$

$$MAPE = \frac{100\%}{n}\sum_{t=1}^{n}\left|\frac{Y_t - \hat{Y}_t}{Y_t}\right|$$

where Y_t represents the actual value, $\hat{Y}_t$ represents the predicted value, and n represents the number of observations in the testing dataset. Lower RMSE, MAE, and MAPE values indicate better forecasting performance. The model with the lowest error values was interpreted as the most suitable forecasting approach for capturing ETF market dynamics.

EMPIRICAL RESULTS

Market Behavior and Data Characteristics

Before evaluating the forecasting models, it is important to examine the statistical behavior of the ETF market data used in this study. Figure 1 presents the historical closing prices of the Invesco QQQ ETF over the observation period from 2019 to 2024. The price trajectory shows a persistent upward trend reflecting the long-term expansion of technology-oriented equities represented in the Nasdaq-100 index. Despite this general growth pattern, the series also exhibits periods of heightened volatility associated with major macroeconomic events, including the market disruption during the COVID-19 pandemic in early 2020 and the global market correction observed in 2022. These fluctuations illustrate the sensitivity of ETF prices to macroeconomic uncertainty, monetary policy expectations, and shifts in investor sentiment.

In contrast to the relatively smooth price trajectory, Figure 2 shows that trading volume exhibits substantially greater variability. The volume series is characterized by irregular spikes that often coincide with periods of significant market activity or the arrival of information. Such spikes frequently emerge during price corrections or major economic announcements, suggesting intensified investor participation during periods of heightened uncertainty. These observations are consistent with financial market microstructure theory, which interprets trading volume as an indicator of information flow and trading intensity.

The differing statistical properties of price and volume series have important implications for forecasting. While price movements tend to follow more structured trends driven by macroeconomic fundamentals, trading volume reflects short-term fluctuations in market sentiment and investor behavior. Consequently, forecasting trading volume is generally more challenging due to its higher volatility and less predictable patterns.

Forecasting Performance of Econometric and Deep Learning Models

The forecasting performance of the ARIMA and LSTM-based models is evaluated using out-of-sample predictions. Table 3 reports the prediction accuracy of each model using three standard error metrics: RMSE, MAE, and MAPE. These metrics provide complementary measures of forecast reliability and enable direct comparison across econometric and deep learning approaches.

The results in Table 3 show that ARIMA recorded the lowest observed forecasting errors among the evaluated models for both target series. For closing-price forecasting, ARIMA produced an RMSE of 3.961, an MAE of 3.119, and a MAPE of 0.82%. Vanilla LSTM was the closest deep learning model, with a MAPE of 1.24%, followed by Stacked LSTM at 1.67% and Bidirectional LSTM at 3.08%.

Table 3. Forecasting Performance Comparison Across Models

Model	Target Variable	RMSE	MAE	MAPE
ARIMA	Closing Price	3.961	3.119	0.82%
	Trading Volume	11,504,271.182	9,217,012.182	19.96%
Vanilla LSTM	Closing Price	5.8404	4.7460	1.24%
	Trading Volume	11,510,951.355	9,233,570.121	20.50%
Bidirectional LSTM	Closing Price	13.2952	11.8898	3.08%
	Trading Volume	11,918,992.945	9,709,287.249	22.14%
Stacked LSTM	Closing Price	6.5020	6.4471	1.67%
	Trading Volume	11,909,650.608	9,659,265.403	21.72%

For trading-volume forecasting, ARIMA also recorded the lowest error values, with an RMSE of 11,504,271.182, an MAE of 9,217,012.182, and a MAPE of 19.96%. However, the substantially higher MAPE for trading volume, compared with the closing-price result, indicates that trading activity remained more difficult to forecast. Therefore, ARIMA should be interpreted as the relatively best-performing model among the evaluated configurations rather than as providing highly accurate volume forecasts.

The numerical differences in Table 3 describe the observed performance of the models on the testing dataset. Because the study did not apply a formal forecast-comparison significance test, the results do not establish statistical superiority between the models.

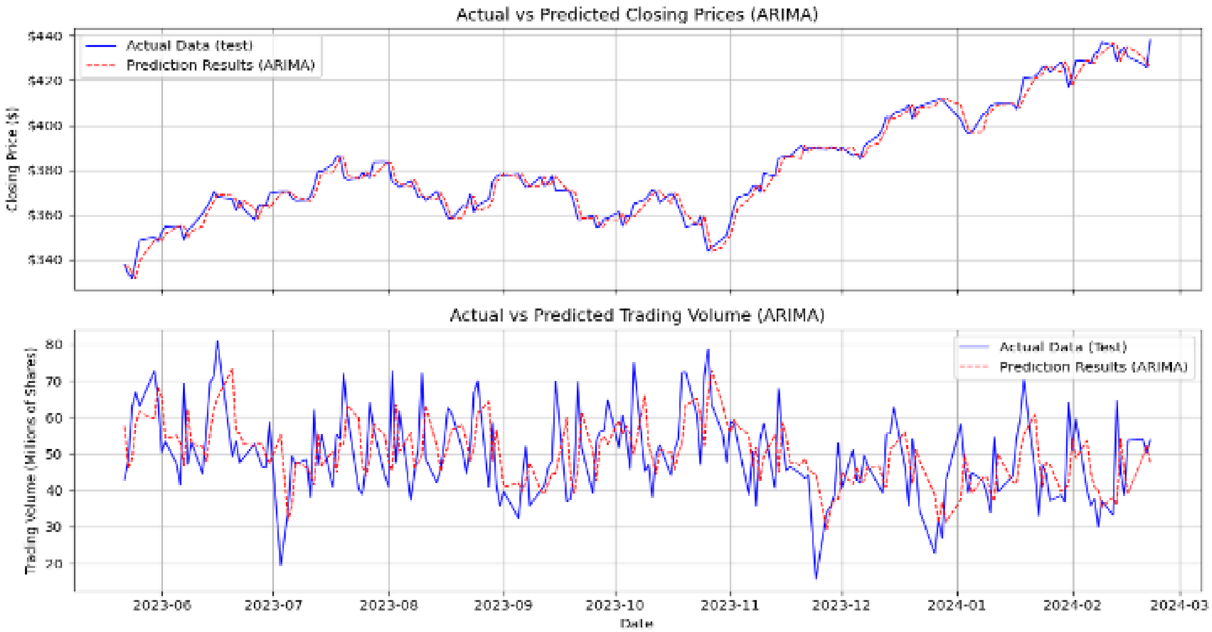

Figure 4. Actual and Predicted ETF Prices and Trading Volume Using the ARIMA Model

Figure 4 compares the actual and predicted closing-price and trading-volume series produced by the ARIMA models. The predicted closing-price series generally follows the direction of the observed price movements, although deviations remain at several points. In contrast, the trading-volume forecasts show larger discrepancies, particularly during abrupt volume spikes. This graphical pattern is consistent with the error metrics in Table 3, which indicate substantially lower relative error for closing price than for trading volume.

Comparative Model Performance and Economic Interpretation

To further evaluate the relative performance of the forecasting approaches, Figures 5 and 6 show graphical comparisons of the predicted ETF closing prices and trading volumes across all models.

Figure 5 shows that the ARIMA predictions were generally closer to the observed closing-price series than the predictions produced by the LSTM architectures. Among the deep learning models, Vanilla LSTM produced the lowest closing-price errors. Bidirectional LSTM and Stacked LSTM showed larger deviations, particularly during periods of more rapid price movement.

Figure 6 shows that all evaluated models experienced greater difficulty in forecasting trading volume. Abrupt fluctuations and volume spikes were not closely reproduced by the model predictions. ARIMA recorded the lowest numerical error, while Vanilla LSTM was the best-performing LSTM architecture. Nevertheless, the MAPE values of all models remained approximately 20% or higher, indicating limited precision for practical volume forecasting.

From an economic perspective, these results carry important implications. In highly liquid financial markets such as the QQQ ETF, new information is rapidly incorporated into asset prices, resulting in relatively structured short-term price dynamics. Under such conditions, classical econometric models that capture linear temporal dependencies may remain highly effective forecasting tools. Conversely, more complex neural network architectures may introduce additional model variance without generating meaningful improvements in predictive accuracy, particularly when the available dataset is limited.

Overall, the empirical evidence suggests that model interpretability and parsimony remain valuable considerations in financial forecasting applications. While deep learning approaches offer flexible nonlinear modeling capabilities, traditional econometric methods such as ARIMA continue to provide robust and reliable predictions in highly liquid ETF markets.

Discussion and Implications

The empirical findings of this study provide several insights into the forecasting behavior of highly liquid ETF markets. Despite the rapid advancement of machine learning techniques in financial forecasting, the results indicate that the traditional ARIMA model delivers the most accurate predictions for both ETF closing prices and trading volumes within the examined period. This finding suggests that short-term ETF price dynamics remain largely driven by structured temporal dependencies that classical econometric models can effectively capture.

From an economic perspective, the lower observed errors of ARIMA may indicate that the short-term movements in the examined QQQ series were sufficiently represented by a parsimonious linear structure during the selected observation period. However, this interpretation should not be treated as direct evidence of market efficiency or universal linearity. The findings are limited to the selected instrument, observation period, data partition, and model configurations. Different results may emerge for other ETFs, longer forecasting horizons, or periods of extreme market stress.

The results also show that increasing neural-network complexity did not improve forecasting accuracy in the present experiment. Vanilla LSTM produced lower errors than Bidirectional LSTM and Stacked LSTM for both targets. Several factors may have contributed to this result, including the limited number of daily observations, the use of predefined model configurations, and the greater number of trainable parameters in the more complex architectures. These explanations remain interpretive because the study did not conduct extensive hyperparameter optimization, repeated training runs, or formal overfitting diagnostics. Therefore, the findings indicate that the more complex LSTM configurations were less effective under the present experimental settings, rather than demonstrating that such architectures are inherently unsuitable for financial forecasting.

From a practical perspective, the findings have important implications for financial analysts and portfolio managers. In environments characterized by high market liquidity and rapid information diffusion, simpler and more interpretable forecasting models may remain reliable tools for short-term market prediction. Consequently, model interpretability and robustness should remain key considerations when selecting forecasting methods for financial applications.

Figure 5. Comparison of ETF Closing Price Forecasts Across Models

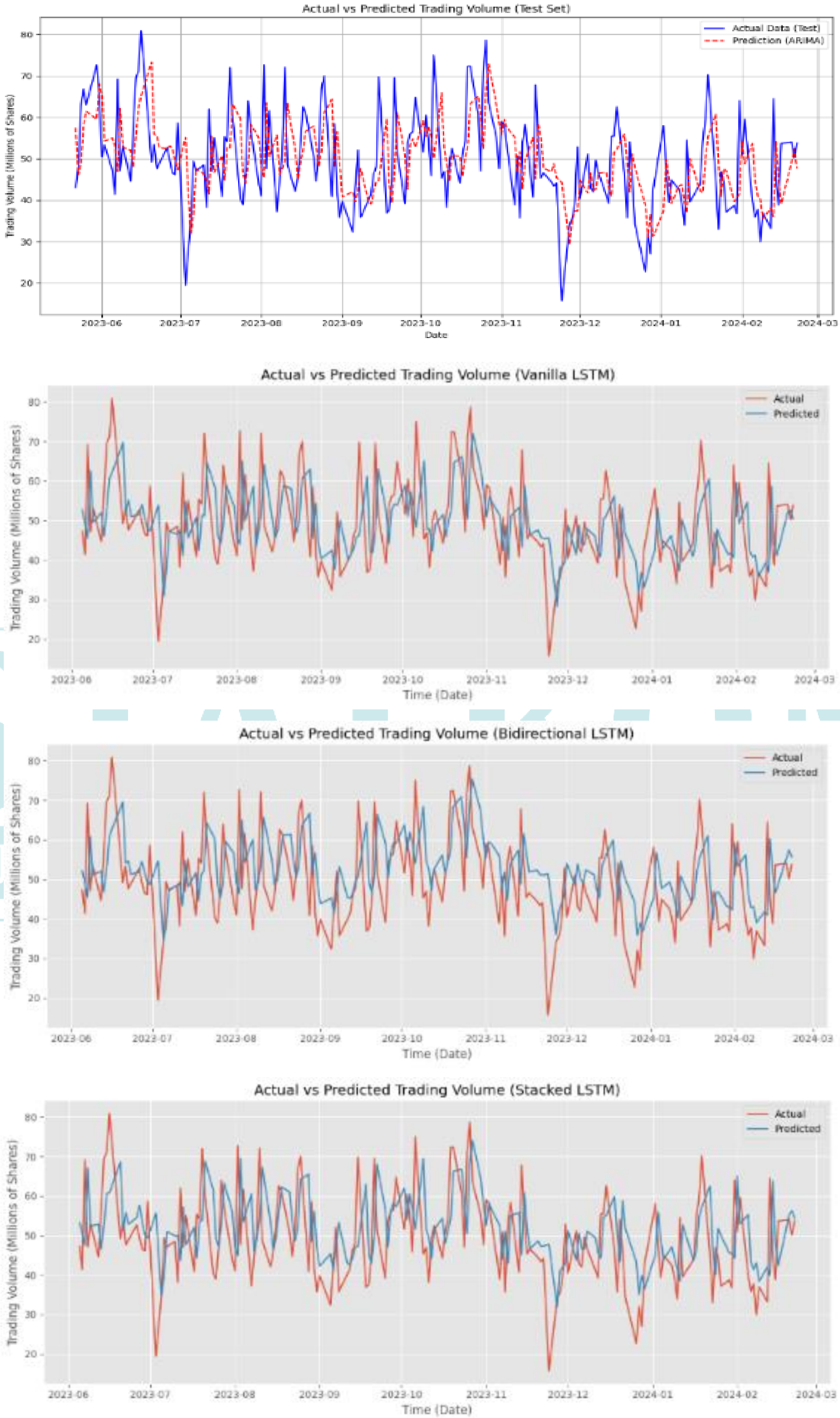

Figure 6. Comparison of ETF Trading Volume Forecasts Across Models

CONCLUSION

Conclusion

This study compared separate ARIMA(7,1,0) models with Vanilla LSTM, Bidirectional LSTM, and Stacked LSTM architectures for forecasting the closing price and trading volume of the Invesco QQQ Trust using daily observations from 2019 to 2024. Among the evaluated configurations, ARIMA recorded the lowest forecasting errors for both targets, with MAPE values of 0.82% for closing price and 19.96% for trading volume. Vanilla LSTM was the most accurate deep learning architecture, whereas Bidirectional and Stacked LSTM produced larger errors under the experimental settings used in this study.

The principal contribution of this study is the evidence that increasing architectural complexity did not improve forecasting accuracy for the examined QQQ series. The results highlight the continuing practical relevance of interpretable and parsimonious econometric models when they adequately represent the temporal structure of the selected data. However, the findings should not be interpreted as proof that ARIMA is universally superior to deep learning. The conclusion is limited to the Invesco QQQ Trust, the 2019–2024 observation period, the selected chronological data partition, and the model configurations evaluated in this study. In addition, the trading-volume MAPE of 19.96% indicates that volume forecasting remained considerably less precise than closing-price forecasting.

Suggestions

This study was limited to one ETF, one chronological train-validation-test partition, and predefined ARIMA and LSTM configurations. It did not employ walk-forward validation, formal forecast-significance testing, extensive hyperparameter optimization, or additional market and macroeconomic predictors. Future research should evaluate multiple ETFs, repeated or walk-forward forecasting schemes, alternative forecasting horizons, and periods of extreme market stress, such as abrupt liquidity disruptions or flash-crash-like events. Further studies may also incorporate volatility indices, interest rates, macroeconomic indicators, and market sentiment to determine whether nonlinear or hybrid models become more advantageous under more turbulent market conditions.

ACKNOWLEDGMENTS

The authors would like to express their sincere gratitude to Universitas Muhammadiyah Surakarta for its continuous support of research and international scientific publications. The authors also acknowledge the Directorate of Research, Community Service, Publication, and Intellectual Property Center (DRPPS), Universitas Muhammadiyah Surakarta, for its publication support. The authors are grateful to the anonymous reviewers for their valuable comments and constructive suggestions, which have significantly improved the quality of this manuscript.

BIBLIOGRAPHY

Ajay, J., Kavitha, R., Ashwin, V. S., & Dhanya, M. (2025). A comparative analysis of ETF performance using machine learning algorithms and traditional models. *In Lecture notes in networks and systems* (pp. 120–128). https://doi.org/10.1007/978-3-032-06691-6_13

Álvarez, P., Muñoz-Palma, M., Bernal, M., García-Gastélum, T., & López-Parra, P. (2025). Intelligent multicriteria classification of ETFs. *In Lecture notes in computer science* (pp. 496–502). https://doi.org/10.1007/978-3-032-10489-2_42

Amankwah, E., Moradi-Motlagh, A., & Valadkhani, A. (2025). Asymmetric responses of the sectoral ETFs to market extremes: evidence from the left and right tails. *Studies in Economics and Finance*, 43(2), 334–353. https://doi.org/10.1108/sef-06-2025-0381

Apriantoro, M. S., Rosadi, R. D. P., Ramdhani, A. C., & Andriyani, N. (2024). Shaping the Future of Environmental Economics: A bibliometric review of current trends and future directions. *International Journal of Energy Economics and Policy*, 14(3), 549–559. https://doi.org/10.32479/ijeep.15502

Bashir, U., Singh, K., Mansotra, V., Khanday, A. M. U. D., & Neshat, M. (2026). EvoBagNet: a decomposition-driven bagging ensemble with evolutionary hyperparameter optimization for robust stock price prediction. *Neural Computing and Applications*, 38(3). https://doi.org/10.1007/s00521-025-11789-z

Chen, Y. (2025). Intelligent carbon price prediction system based on improved secondary decomposition and multi-head attention bidirectional long short-term memory (BiLSTM) model. *Scientific Reports*, 16(1), 1803. https://doi.org/10.1038/s41598-025-31325-2

Chen, Z., Chen, H., Liu, Y., & Shi, Y. (2026). Multi-label financial statement fraud detection based on long short-term memory and multilayer perceptron hybrid model. *Engineering Applications of Artificial Intelligence*, 170, 114188. https://doi.org/10.1016/j.engappai.2026.114188

Chichernea, D. C., Petkevich, A., & Wang, K. (2026). Short selling ETFs and market performance. *The Journal of Financial Research*. https://doi.org/10.1111/jfir.70050

Divelgama, B., Nyarko, N. A., Aryee, N. S. D., Shirvani, A., & Rachev, S. T. (2026). Performance and risk analytics of Asian Exchange-Traded Funds. *Journal of Risk and Financial Management*, 19(1), 69. https://doi.org/10.3390/jrfm19010069

Frimpong, V., & Mamuti, A. (2026). Opaque price control and algorithmic authority in financial markets. *Encyclopedia*, 6(1), 19. https://doi.org/10.3390/encyclopedia6010019

Geng, R., Zhang, H., Gao, Y., & Yuan, G. (2025). Decoding global economic dynamic: A graph-based examination of contemporary ETF markets. *Chaos Solitons & Fractals*, 201, 117313. https://doi.org/10.1016/j.chaos.2025.117313

Lee, C., Sah, A., & Huang, H. (2026). Impact of investor sentiment on stock market using sentiment estimation and multi-head attention LSTM network. *Journal of the Chinese Institute of Engineers*, 1–14. https://doi.org/10.1080/02533839.2026.2619709

Li, C. (2025). Real Estate Market Trend Prediction: A Time Series Model Based on LSTMNeural Networks and Its Application in Financial Risks. *Proceedings of the 2025 2nd International Conference on Digital Economy and Computer Science*, 364–368. https://doi.org/10.1145/3785706.3785764

Maryam, M., Anggoro, D. A., Tika, M. F., & Kusumawati, F. C. (2022). An intelligent hybrid model using artificial neural networks and particle swarm optimization technique for financial crisis prediction. *Pakistan Journal of Statistics and Operation Research*, 1015–1025. https://doi.org/10.18187/pjsor.v18i4.3927

Mo, F. (2025). Machine Learning and Deep Learning in Quantitative Finance. *Proceedings of the 2025 2nd International Conference on Digital Economy and Computer Science*, 333–339. https://doi.org/10.1145/3785706.3785759

Naifar, N. (2026). Systemic tail dependence in disruptive technology ETFs & crypto assets: A partial correlation network. *Research in International Business and Finance*, 84, 103316. https://doi.org/10.1016/j.ribaf.2026.103316

Novotny, J., & Hajek, P. (2026). A hybrid adaptive trading strategy integrating investor sentiment for precious metal ETFs. *Financial Innovation*, 12(1). https://doi.org/10.1186/s40854-026-00911-2

Nuryakin, N., Widayanti, R., Damayanti, R., & Susanto, N. (2021). The importance of market information accessibility to enhancing SMEs Indonesian superior financial performance. *International Journal of Business Innovation and Research*, 25(1), 1. https://doi.org/10.1504/ijbir.2021.115010

Ozcelebi, O., McIver, R., & Kang, S. H. (2025). The dynamics of frequency connectedness between technology ETFs and uncertainty indices under extreme market conditions. *Financial Innovation*, 11(1). https://doi.org/10.1186/s40854-025-00760-5

Rios-Vazquez, S., & Portela-Maseda, M. (2026). Fintech empirical studies: a state-of-the-art systematic literature review. *Financial Innovation*, 12(1). https://doi.org/10.1186/s40854-026-00918-9

Sahu, H. K., Parida, I., Bhandari, A., Jena, N. R., & Vijay, M. (2026). A Random Walk down Cross-Asset Networks: A Deep Learning Tour of Volatility Transmission between Digital and Conventional Markets. *Asia-Pacific Financial Markets*. https://doi.org/10.1007/s10690-025-09588-6

Silfiani, M., Lumintang, I. A., & Winda, R. L. (2024). Perbandingan Beberapa Metode Univariat Time Series pada Peramalan Curah Hujan. *Jurnal Statistika Dan Komputasi*, 3(1), 22–31. https://doi.org/10.32665/statkom.v3i1.2730

Solanki, N., & Jha, M. (2025). From Time Series to Images: A GAF-RESNET-LSTM framework for asset selection and portfolio optimization. *IEEE Access*, 14, 9490–9505. https://doi.org/10.1109/access.2025.3648467

Su, Y., & Wang, M. C. (2026). Correlated bivariate time series forecasting using long short-term memory network: an AutoML approach. *International Journal of Data Science and Analytics*, 22(1). https://doi.org/10.1007/s41060-026-01036-6

Tsoku, J. T., & Makatjane, K. (2026). Deep learning-based pairs trading: real-time forecasting of co-integrated cryptocurrency pairs. *Frontiers in Applied Mathematics and Statistics*, 12. https://doi.org/10.3389/fams.2026.1749337

Utami, D. E., & Irawati, Z. (2021). The role of the financial and macroeconomy industry on the development of the Sukuk (Sharia Compliant Bonds) market: The case of Indonesia.

Academic Journal of Interdisciplinary Studies, 10(4), 225. https://doi.org/10.36941/ajis-2021-0112

Xu, H., Yuan, X., Jin, L., Long, J., & Xu, G. (2025). Ascertaining price formation in financial markets with machine learning: Evidence from Chinese stocks. *Pacific-Basin Finance Journal*, 96, 103029. https://doi.org/10.1016/j.pacfin.2025.103029

Zeng, L., Huang, J., & Lin, X. (2026). LSTM-augmented vine copula modelling for energy-finance contagion analysis. *Scientific Reports*, 16(1), 5358. https://doi.org/10.1038/s41598-026-37150-5