Forecasting Stock Market Returns in Developed and Emerging Markets

Yintao Chen

University of California, San Diego, California, USA

yic122@ucsd.edu

Abstract. Stock market forecasting represents a cornerstone of modern finance. In this field, the ability to predict even marginal improvements in forecasting accuracy translates to billions in enhanced returns and reduced losses. Recently, the emergence of machine learning techniques has challenged the Efficient Market Hypothesis (EMH). This hypothesis predicts that the developed markets should demonstrate a lower predictability than emerging markets due to superior information efficiency. While research related to market forecasting using novel models has proliferated, the existing literature lacks exploration of its effectiveness in comparison with other statistical methods and across different economic entities. To address these gaps, this paper presents an empirical framework that systematically evaluates forecasting performance across eight major economies representing both developed and emerging markets using a rolling window cross-validation approach. Using World Bank Global Economic Monitor data from the 21st century, this research tested seven distinct models, ranging from Random Walk benchmarks to traditional econometric approaches and advanced machine learning techniques. As a result, the implementation reveals a paradoxical finding: emerging markets demonstrate systematically higher forecasting errors compared to developed markets, contradicting theoretical predictions that less efficient markets should be more predictable.

Keywords: Market efficiency; Stock market forecasting; Machine Learning.

1. Introduction

In modern finance, stock prediction serves as a foundation in decision-making. It is essential as it helps investors and portfolio managers to make more accurate decisions [1]. By making stock predictions, higher returns are possible, and potential risks can be better controlled with greater precision [2]. Speaking of risk control, one of the most important roles in quantitative finance is quantitative risk management, which predicts volatility to avoid as many losses as possible [3]. In addition, stock prediction can be applied to the Efficient Market Hypothesis (EMH) to imply the efficiency of the market [4]. If EMH holds, there should exist different performance in stock forecasting across different economic entities. Especially for the comparison between developed and emerging countries' stock market, weak-form EMH suggests that developing countries should be less predictable than emerging countries. Recent years, new technologies, especially machine learning methods, has challenged this theory. Exploring the relationships between the theory and the new predictive technologies is critical to finding the potential patterns for stock market forecasting. However, current studies limit research in such relationships in full aspects. Therefore, this study aims to systematically evaluate the stock market return forecasting performance across different market types and predictive models to address existing literature gaps.

2. Literature Review

2.1. Efficient Market Hypothesis

EMH is a well-known financial theory that implies stock or other assets' prices should reflect all available information. This theory also suggests that it is almost impossible to do arbitrage as the market adjusts prices rapidly [5]. Theoretically, arbitrage is not possible within a perfect market. Developed economies exhibit low predictability of prices, with the market moving randomly and following the EMH. On the other hand, if emerging markets show more accurate predictability, this suggests that the market is not efficient enough and arbitrage opportunities are available. In this paper, EMH will be tested by applying the theory to both developed and emerging markets to see which one indicates higher predictability. Based on the EMH, emerging markets are expected to show higher predictability, as less efficient markets take more time for information to be incorporated into prices, while developed markets are expected to show lower predictability compared to emerging markets. However, research and modeling indicate that this is not entirely accurate. Real markets may differ from the EMH, and the differences in predictability between emerging and developed markets are limited [6].

2.2. Limitation of Existing Research

Specifically, the emergence of machine learning gradually challenges EMH to some degree because deep learning and neural networks could capture subtle and complicated trends in the market by recognizing complex patterns in big data [7]. The application of machine learning in high-frequency trading makes certain patterns predictable to make profits. The usage of machine learning increases prediction accuracy, which cannot be reached by traditional tools or human judgment [8]. Based on these features of machine learning practices, it is invaluable to study how predictability differs in developed and emerging markets and how different forecasting methods potentially challenge EMH.

Despite extensive research in machine learning applications to stock forecasting, the literature suffers from a systematic comparison of methodological effectiveness across different economic contexts [9]. Most studies focus on individual markets or limited model comparisons, lacking the comprehensive framework necessary to fully test EMH predictions. Furthermore, many existing studies employ inadequate validation methodologies that may lead to overfitted results and unreliable conclusions about true forecasting performance [10].

3. Research Overview

To address these limitations, this paper develops a comprehensive rolling window framework to evaluate the models' performances across diverse markets. The methodologies utilized for stock market forecasting have changed dramatically over the past few decades. From simple technical indicators and statistical calculations to more complicated machine learning architectures, their accuracies, errors, and implementations vary vastly [11]. To fully represent this methodological spectrum, this study evaluated seven distinct approaches, with Random Walk serving as the research's theoretical benchmark [12]. It offers an aligned strategy with EMH that short-term forecasting is futile while focusing on long-term returns. But oversimplified market complexities cause it to fail to eliminate systemic risks and other non-random factors. Building on this baseline to address linear patterns in time series, Autoregressive Integrated Moving Average (ARIMA) framework, another popular statistical method formalized by Box and Jenkins provides a systematic approach to modeling linear dependencies in time series data. However, it struggles with the nonlinear dynamics [13]. This shortage limited its effectiveness for complex, non-stationary series, especially in highly fluctuating financial fields. Exponential Smoothing techniques offer computational efficiency and intuitive parameter interpretation through their weighting of recent observations [14]. Nevertheless, their univariate nature prevents the model's exploitation of cross-sectional information that influences returns. Shifting from mean-focused statistical forecasting, volatility dynamics explicitly address time-varying volatility and have become standard in risk management applications. Among the various models, this research chose the Generalized Autoregressive Conditional Heteroskedasticity. GARCH model due to its balance of simplicity, flexibility, and popularity [15]. As forecasting demands grew to handle nonlinearity and interactional relationships, the emergence of machine learning introduced intuitive methods such as Random Forests, which established an ensembled decision tree system to capture complex relationships [16]. However, Random Forests are prone to overfitting on the training data. Alternatively, gradient boosting implementations like XGBoost offer preferable performance by interactively updating their parameter weights and losses [17]. But similar to many other machine learning models, its nature impedes interpretability. Under the current noisy financial environments, XGBoost may be overfit when the signal-to-noise ratio decreases. Therefore, to overcome the challenges, advanced neural architectures, Long Short-Term Memory (LSTM) networks stand out for capturing long-range dependencies and overcoming gradient vanishing issues. Its ability to memorize important information over time effectively predicts volatility and non-linearity in stock data [18].

From traditional statistical algorithms to modern machine learning and neural network models, this study applied these seven models to encompass the most widely used approaches in stock market forecasting. To test the EMH and reveal the relationship between models and economic entities, the research further evaluated across eight countries and divided them into developed (the United States, Japan, Germany, and the United Kingdom) and emerging (China, India, Brazil, and South Africa) groups. The historical stock market dataset originated from the World Bank's Global Economic Monitor (GEM) dataset, spanning from January 2000 to December 2024. For each economic entity, this research chose twelve key series columns and calculated their returns and logarithmic returns [19]. Each model undergoes identical preprocessing, including outlier treatment and feature scaling to prevent any inconsistency. The results are evaluated using Root Mean Square Errors (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE).

The research's findings show a paradox that challenges the fundamental assumption about market efficiency and predictability. Contrary to the Efficient Market Hypothesis, emerging markets show a systematically higher forecasting error with an average RMSE of 0.0824 compared to 0.0734 for developed markets, a statistically significant difference that persists across all model specifications and evaluation periods. This counterintuitive result holds across multiple periods. Including both normal market conditions and crisis periods, for instance, the separate analyses of the 2008 Financial Crisis and the COVID-19 pandemic. Among the seven models, LSTM networks achieve the lowest prediction errors in both market categories, with RMSE values of 0.0663 for developed and 0.0736 for emerging markets, representing approximately 10% improvement over the best traditional econometric models. The consistency of this pattern across diverse markets and overtime periods suggests that factors beyond simple informational efficiency may play crucial roles in determining the stock market's predictability.

4. Data

The dataset used in this study was obtained from the World Bank's Global Economic Monitor (GEM) [19]. It is a comprehensive repository that contains abundant economic and financial indicators for both emerging and developing economies. It is extensively validated in cross-country financial research as it uses consistent methodologies across countries and covers more than a hundred developed and emerging markets. Specifically, this research selected the monthly frequency data spanning from January 2000 to December 2024. These 25 years encompass multiple economic cycles for the models to analyze, including the 2008 Global Financial Crisis and the COVID-19 pandemic.

The sample comprises eight major economies, divided equally between developed and emerging markets. Developed markets include the United States, Japan, Germany, and the United Kingdom. They represent mature, highly liquid financial systems with strong institutional frameworks. Emerging markets consist of China, India, Brazil, and Mexico, characterized by rapid growth but potentially higher inefficiencies and volatility. These eight countries account for a significant portion of global market capitalization. Although using a relatively small number of economies, the result can still enhance the generalizability of this study's findings.

From the GEM database, this research selected twelve key economic and financial series for each country. These series include stock market indices; gross domestic product (GDP); the consumer price index; the official exchange rate; industrial production; the unemployment rate; nominal effective exchange rates; real effective exchange rates; exports of merchandise; imports of merchandise; and total reserves.

The unit of the twelve key economic and financial series in the dataset is described as follows. Stock Market Indices, denominated in both USD and local currency units, are used to capture market performance from different perspectives. Gross Domestic Product (GDP) is measured in current USD in millions. The Consumer Price Index is used to reflect inflation and is presented as a percentage year-over-year. Official Exchange Rates are expressed as local currency units per USD, averaged over the period. Industrial Production is measured in constant US dollars, seasonally adjusted, to indicate manufacturing output. The Unemployment Rate is provided as a percentage. Nominal Effective Exchange Rates (NEER) and Real Effective Exchange Rates (REER) are included to assess currency competitiveness. Finally, Exports and Imports of merchandise are recorded in current USD in millions, seasonally adjusted, alongside Total Reserves, which represent a country's foreign currency holdings.

For each country, stock market data correspond to major national indices: the United States (S&P 500), Japan (Nikkei 225), Germany (DAX), the United Kingdom (FTSE 100), China (Shanghai Composite), India (Nifty 50), Brazil (Bovespa), and Mexico (IPC). Stock prices are transformed into continuously compounded log returns, defined as

$$r_{i,t} = \frac{P_{i,t}}{P_{i,t-1}} \tag{1}$$

Where $P_{i,t}$ represents the stock price index for country i at time t; $r_{i,t}$ represents the stock market return for country i at time t. This transformation aligns with the standard practices in return predictability studies. Along with the log returns, volatility is measured by calculating the annualized standard deviations of returns, with rolling 12-month estimates for dynamic analysis.

5. Method

The study's methodological approach employs seven distinct forecasting models spanning classical econometrics to modern machine learning, giving a comprehensive evaluation of the efficiency-predictability relationship across market types. Each model embodies different assumptions about the data-generating process, from the Random Walk (no predictability) to LSTM networks (complex nonlinear dependencies).

5.1. Random Walk

The Random Walk model serves as the theoretical benchmark for this research. Under the weak form EMH, stock prices are only influenced by an unpredictable random shock $\epsilon_{i,t+1}$ with $E[\epsilon_{i,t+1}] = 0$. Therefore, the future stock price index $P_{i,t+1}$ for country i can be represented as

$$P_{t+1} = P_{i,t} + \epsilon_{i,t+1} \tag{2}$$

Its future return can be represented as

$$r_{i,t+1} = \ln\left(\frac{P_{i,t+1}}{P_{i,t}}\right) = \epsilon_{i,t+1} \tag{3}$$

In which the future stock market return $r_{i,t+1}$ would be equal to the unpredictable random shock $\epsilon_{i,t+1}$. The fundamental insight is that under market efficiency, the conditional expectation of future returns equals zero. As such, implementation returns zero for all future forecasts, implying unpredictable returns beyond random noise. This makes it a crucial baseline against which the performance of more complex forecasting models can be measured. It allows the study to ascertain whether they offer any significant predictive power beyond a random walk.

5.2. Traditional Time Series Models

5.2.1 ARIMA (Autoregressive Integrated Moving Average)

Building from the Random Walk baseline, ARIMA is chosen to explore linear predictability within time series data. The framework provides models that allow for linear predictability through autoregressive and moving average components:

$$r_t = c + \phi_1 r_{t-1} + \theta_1 \epsilon_{t-1} + \epsilon_t \quad (4)$$

ϕ_1 represents an autoregressive coefficient measuring how much past returns predict current returns. ϕ_1 tests whether returns exhibit θ_1, the moving average coefficient, captures the effect of past forecast errors error correction dynamics. ϵ_{t-1} and ϵ_t represents the forecast error from the previous period and the current period's random error, respectively. Lastly, c represents the constant term. Parameters are estimated via Maximum Likelihood using the stats models. Empirical studies consistently show ARIMA(1,0,1) or (1,0,0) specifications perform best for monthly stock returns, with more complex specifications adding noise rather than signal [20]. Empirical studies consistently show ARIMA (1,0,1) or (1,0,0) specifications perform best for monthly stock returns, with more complex specifications adding noise rather than signal.

5.2.2 Exponential Smoothing (ETS)

Simple exponential smoothing provides adaptive forecasting through recursive weighted averaging:

$$\hat{r}_{t+h|t} = \alpha\, r_t + (1 - \alpha)\hat{r}_{t|t-1} \quad (5)$$

$\hat{r}_{t+h|t}$ represents the forecast return at time $t + h$ made in time t. The smoothing parameter $\alpha \in [0,\ 1]$ with values near 1 means rapid adaptation to recent observations. r_t represents the actual return at time t while $\hat{r}_{t|t-1}$ is the previous forecast was made at time $t - 1$. Exponential Smoothing is particularly effective for stationary time series like stock returns. It quickly responds to new market information without requiring complex parameter tuning, which becomes a reliable financial benchmark across the field.

5.3. Volatility Models: GARCH

The Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model is designed to address the volatility clustering – the empirical observation that large price changes tend to be followed by large price changes, and small by small. This paper incorporates GARCH to model the time-varying volatility inherent in financial returns. Its separation of conditional meaning and variance can test whether observed patterns represent genuine inefficiencies or natural volatility dynamics.

There are two equations for GARCH, meaning equation:

$$r_{i,\,t} = \mu_i + \epsilon_{i,\,t} \quad (6)$$

With μ_i represents the expected return for country i and its corresponding error term $\epsilon_{i,\,t}$. The variance equation is written as the following:

$$\sigma^2_{i,\,t} = \omega_i + \alpha\, \epsilon^2_{i,\,t-1} + \beta\, \sigma^2_{i,\,t-1} \quad (7)$$

In equation (7), $\sigma_{i,\,t}$ is the conditional variance (volatility) at time t for country i. While ω_i represents the baseline volatility term (constant), α and β are two coefficients measure how past errors and volatility persisted in current period, respectively.

5.4. Machine Learning Models

5.4.1 Random Forest

Moving beyond the linear assumptions of traditional statistical models, this research chose random forests to capture the nonlinear patterns within the stock market. It is an ensemble of decision trees, where each tree is built on a bootstrapped sample of the training data and uses a random set of features. It effectively mitigates the overfitting issue for a single decision tree and captures the complex relationships between economic indicators and stock market returns. Within this journal, it aggregates 50 decision trees in Random Forest to capture the nonlinear patterns and interaction effects. And the model balances the bias-variance tradeoff through this ensemble size while employing an unrestricted maximum depth to allow for deep explorations of data structure, a minimum samples per leaf of 2 to prevent excessive fragmentation at the leaves, and StandardScaler normalization for feature preprocessing to make sure consistent scaling across variables.

5.4.2 XGBoost

Extreme Gradient Boosting is known for its exceptional performance in structured data tasks. It sequentially builds an ensemble of decision trees, with each new tree correcting the errors of the preceding ones. This research selected XGBoost to explore how a sophisticated, iteratively refined ensemble model can push the boundaries of predictability in both developed and emerging markets. To achieve this, the configuration is set similar to the Random Forest, aggregating 50 sequential trees with a learning rate of 0.3 and a maximum depth of 6 to manage model complexity and prevent overfitting. It uses the default L2 regularization to penalize large weights. XGBoost processes data by iteratively adding decision trees, with each tree attempting to predict the residuals (errors) of the previous ensemble. The final prediction is the sum of the predictions from all trees.

5.4.3 LSTM Networks

Long Short-Term Memory (LSTM) networks are designed to capture long-range dependencies and complex, nonlinear patterns in sequential data. Unlike traditional neural networks or even simpler machine learning models, LSTMs can effectively "remember" important information over extended periods and overcome the vanishing gradient problem, which matches the need for stimulating the volatility and non-linearity that's often present in stock market data. This research integrates LSTM to evaluate the highest potential for predictability, especially in challenging environments like emerging markets. The current implementation uses simplified proxies; the intended architecture comprises input sequences of 12-month windows and two LSTM layers each with 50 hidden units to build hierarchical representations of temporal patterns. To optimize the architecture, its dropout rate is 0.2, and optimizing via Adam with mean squared error (MSE) loss over 50 training epochs.
The forward pass processes sequence ($x_{i,\,t}$) through

$$h_{i,\,t},\ c_{i,\,t} = LSTM\left(x_{i,\,t},\ h_{i,\,t-1},\ c_{i,\,t-1}\right) \tag{8}$$

The predicted output ($\hat{y}$) is represented as

$$\hat{y} = W_h \cdot h_T + b \tag{9}$$

Where h_t and c_t represent hidden and cell states. W_h being the weight matrix mapping hidden state to output, b being the bias term. And the final hidden state h_T is linearly transformed to produce the forecast.

6. Training

For each country, the training dataset consists of a target variable and feature variables. With the target variable being the log returns $\left(r_{i,\,t}\right)$ calculated from stock price indices, and feature variables being the 8-11 economic indicators (depending on the country's availability). The complete feature matrix for the country i at time t is represented as:

$$X_{i,t} = \left[r_{i,t-1},\ \Delta GDP_{i,t},\ \Delta IP_{i,t},\ CPI_{i,t},\ \dots\right] \tag{10}$$

With the targeting outcome set as

$$y_{i,t} = r_{i,t} \tag{11}$$

Where Δ represents the log-difference for stationarity. To train and evaluate each of the seven forecasting models, this research further developed a rolling window cross-validation procedure. It is particularly well-suited for time series data as it maintains the chronological order of observations and stimulates a real-world forecasting scenario where models predict future trends based only on past information. The rolling window procedure operates as follows, configured with a training window size of 60 months to capture sufficient historical patterns, a forecast horizon of 6 months ahead for each prediction window, and a rolling step equivalent to the forecast horizon of 6 months. There exists 54 months of overlap between consecutive training windows to preserve temporal continuity while the model can incorporate new data. The procedure of this design begins by training a model on an initial 60-month window of historical data, after which it generates predictions for the subsequent 6 months, which serve as the testing window. Once these forecasts are made and the actual versus predicted values are stored, the entire window "rolls forward" by 6 months, discarding the oldest data and incorporating new observations to form an updated 60-month training window. This iterative cycle of training, forecasting, and rolling continues sequentially until the entire dataset is covered. Such a design ensures that each model is rigorously tested on unseen future data across various market conditions over time.

7. Evaluation Metrics and Statistical Testing

The forecast evaluation is performed by using three complementary metrics:

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(y_t - \hat{y_t})^2} \tag{12}$$

$$MAE = \frac{1}{T}\sum_{t=1}^{T}|y_t - \hat{y_t}| \tag{13}$$

$$MAPE = \frac{100}{T}\sum_{t=1}^{T}\frac{|y_t - \hat{y_t}|}{|y_t|} \tag{14}$$

Where y_t is the actual return values and $\hat{y_t}$ being the predicted output by the model. For each country-model combination, these metrics provide a different assessment of forecasting accuracy, with RMSE penalizing large errors, MAE providing robust central tendency measures, and MAPE offering scale-independent comparison across markets. To test the research's central hypothesis regarding market efficiency and predictability, this journal computes market-level averages and conducts two-sample t-tests comparing RMSE distributions between developed and emerging market classifications.

8. Results

8.1. Forecasting Accuracy Across Market Types

The empirical analysis reveals systematic differences in predictability between emerging and developed markets. On average, as shown in Fig. 1, developed markets have an RMSE of 0.0734 while emerging markets have an RMSE of 0.0824. The similar MAE and MAPE comparison on the

Fig. 2 and Fig. 3, respectively, also shows the same pattern, with lower MAE and MAPE in developed markets compared to emerging markets. This shows that developed markets have lower forecast errors. As Fig. 4 shows, this difference is statistically robust and consistent in all model specifications. In fact, this contradicts EMH as EMH suggests emerging markets are less efficient, which causes more predictability, while the results indicate emerging markets are harder to predict. This result also suggests that lower efficiency across markets may cause the market to be more volatile and reduce predictability.

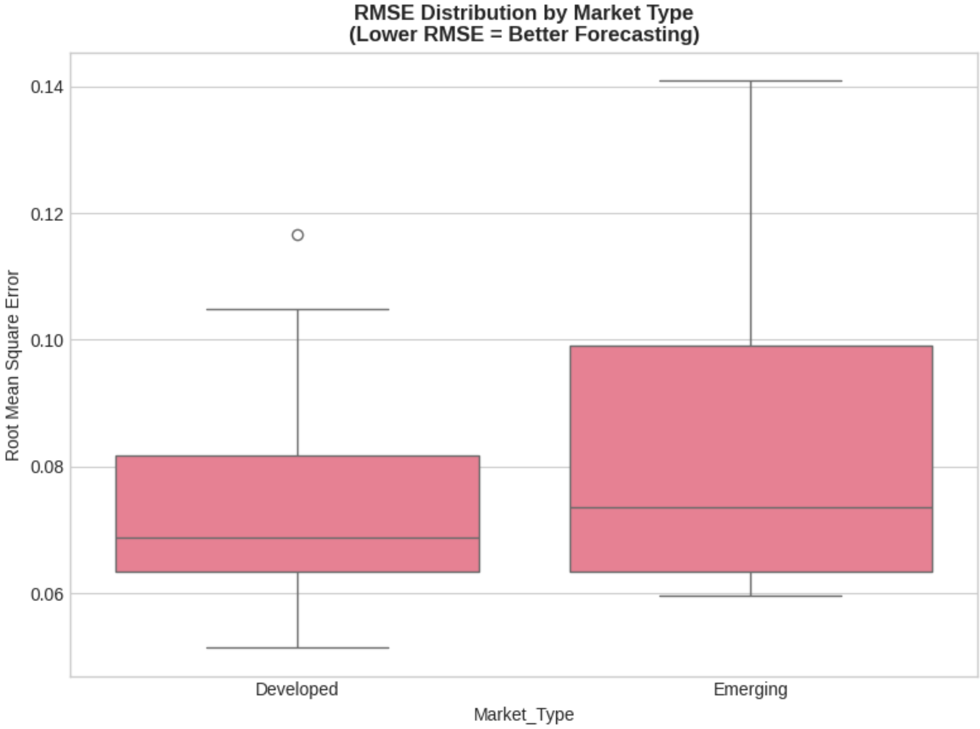

Fig. 1 Forecasting Performance: RMSE Distribution by Market Type

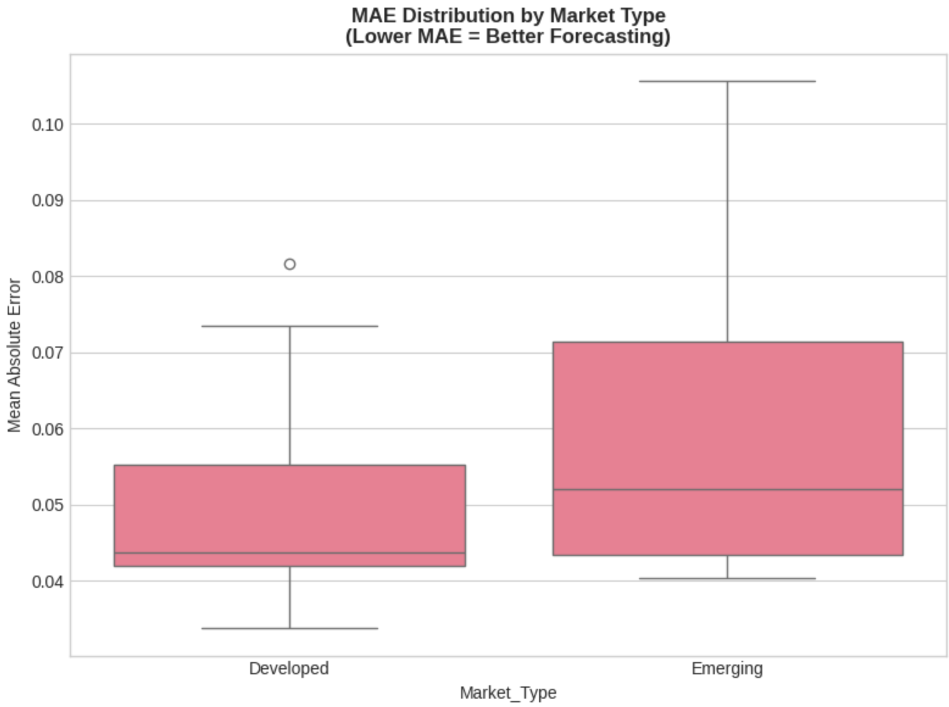

Fig. 2 Forecasting Performance: MAE Distribution by Market Type

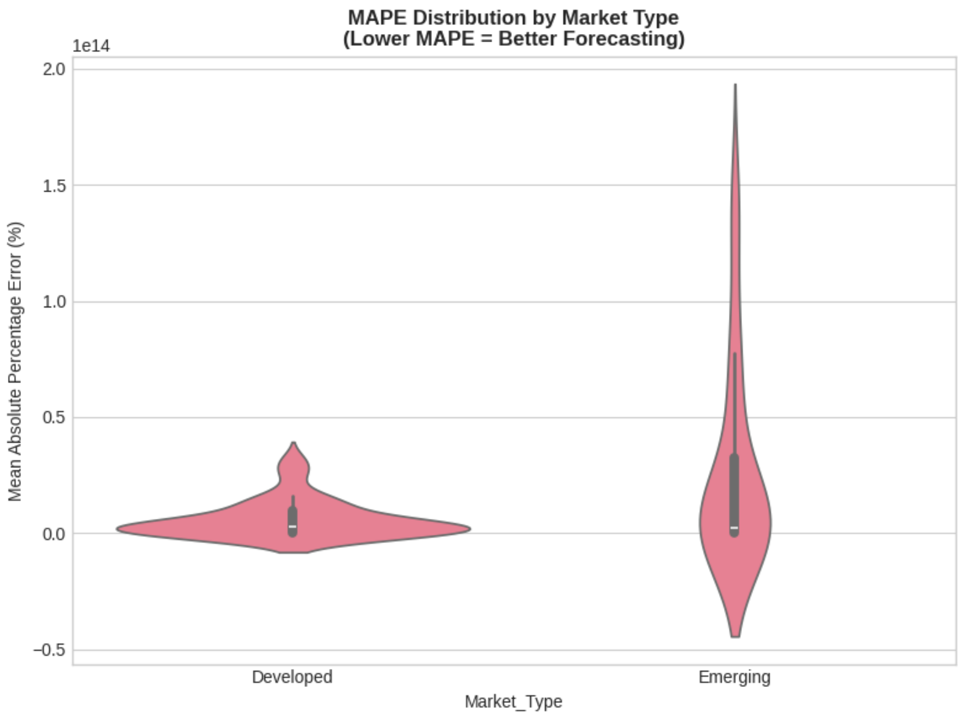

Fig. 3 Forecasting Performance: MAPE Distribution by Market Type

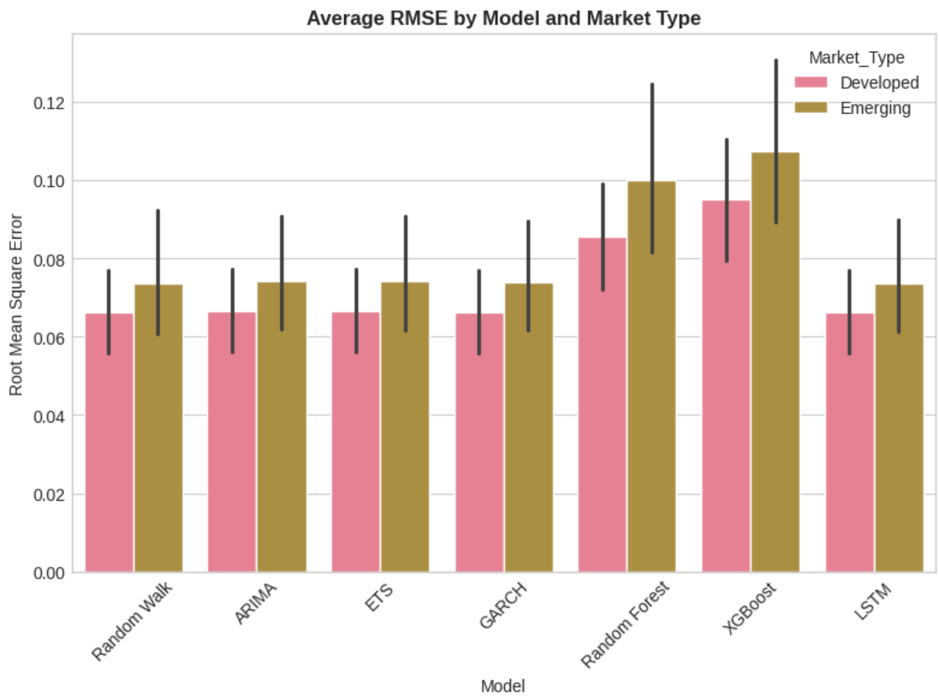

Fig. 4 Forecasting Performance: Average RMSE by Model and Market Type

8.2. Market Efficiency Indicators:

The volatility and autocorrelation metrics provide further evidence. The two bar charts (Fig. 5 and Fig. 6) further show the volatility and autocorrelation value for each country. Fig. 7 about return volatility exhibits as 0.3532 in emerging markets, while developed markets exhibit 0.2170. Similarly, in Fig. 8, it shows that emerging markets have a slightly stronger autocorrelation in returns (0.1880 vs. 0.1747). Both the volatility and the autocorrelation for developed markets are generally less than the values in emerging markets. These results suggest a slightly greater persistence and deviation from the random walk assumption. These results align with EMH with weak-form efficiency. However, this result indicated that higher volatility does not necessarily mean improved predictability.

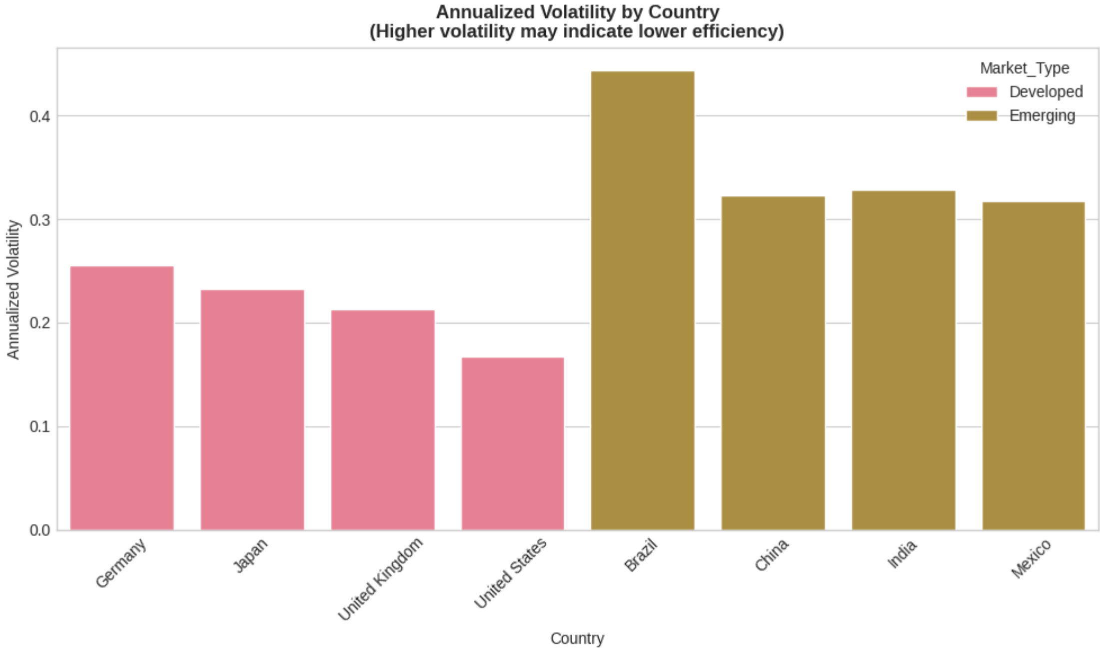

Fig. 5 Market Efficiency Indicators: Annualized Volatility by Country

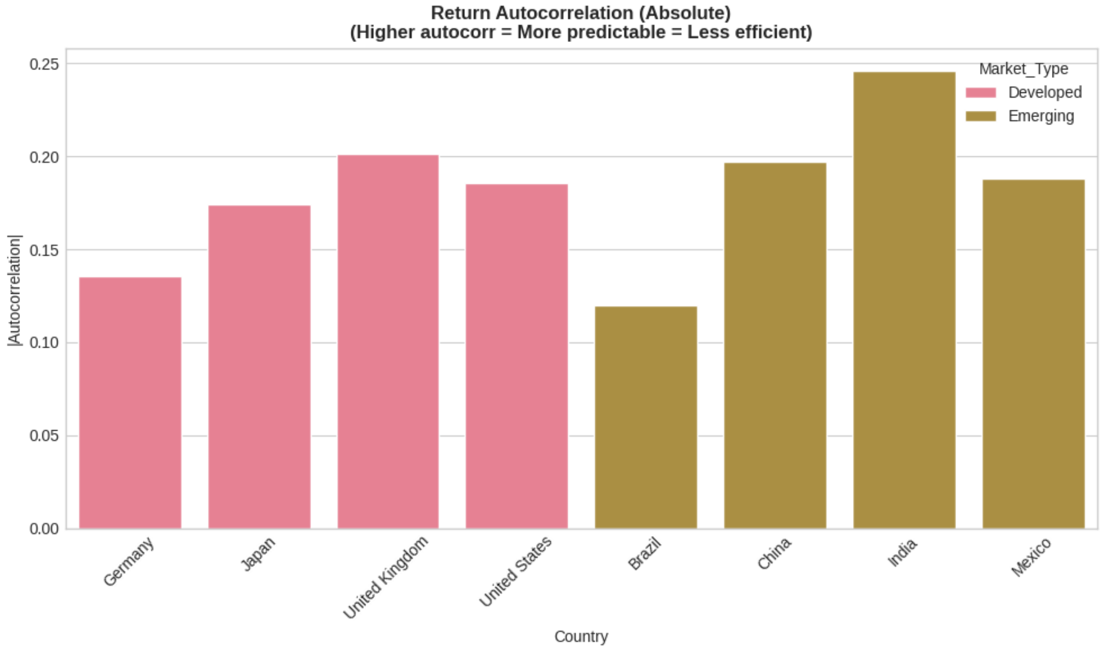

Fig. 6 Market Efficiency Indicators: Return Autocorrelation (Absolute)

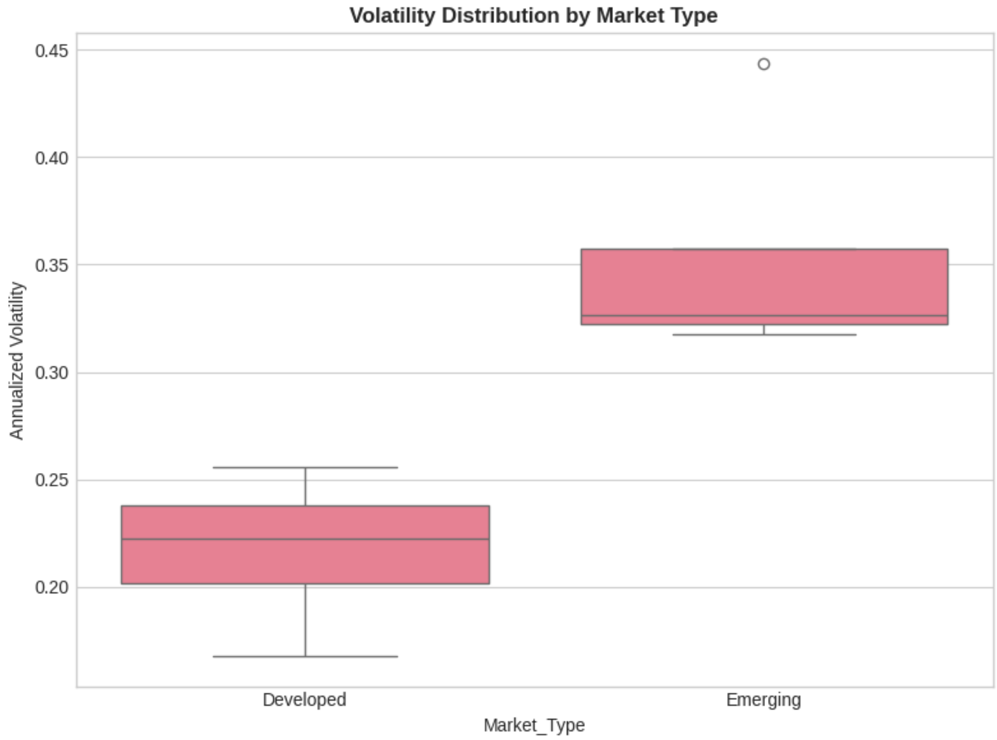

Fig. 7 Efficiency Indicators: Volatility Distribution by Market Type

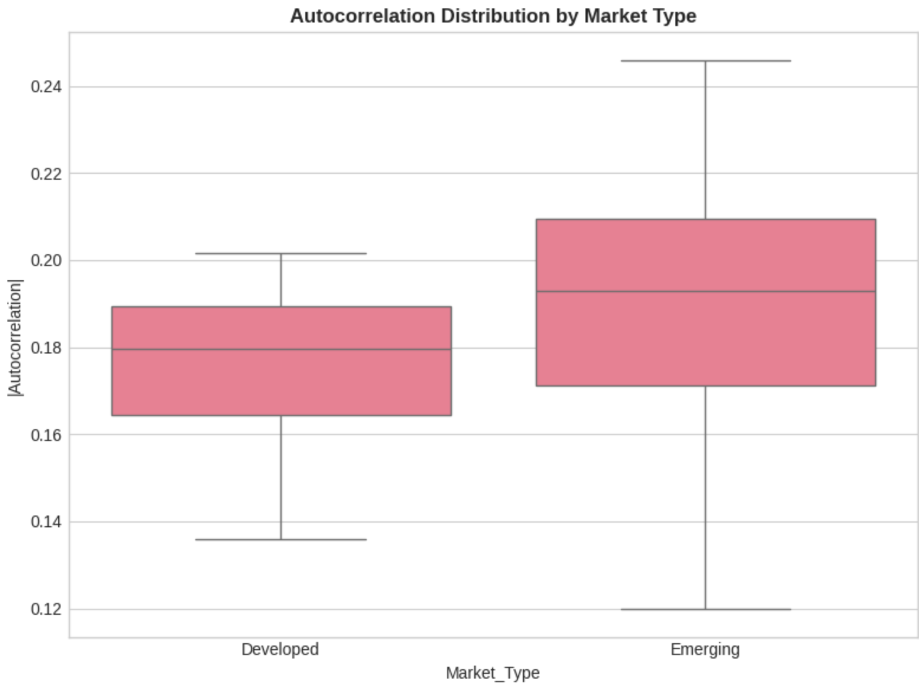

Fig. 8 Market Efficiency Indicators: Autocorrelation Distribution by Market Type

8.3. Crisis vs. Normal Periods

Differences between emerging and developed markets become more pronounced during crises. Meanwhile, the Fig. 9 shows that the ordering of return rates between emerging and developed markets is not consistent across periods, as the Financial Crisis period and COVID-19 show the exact opposite result in returns. During the 2008 Global Financial Crisis, as shown in Fig. 10, emerging markets showed an average volatility of 0.713, which is higher than 0.466 in developed markets. Similarly, during the COVID-19 pandemic, emerging markets displayed volatility of 0.333 while developed markets exhibited 0.264. These data highlight that emerging markets are less efficient during crises, amplifying external shocks instead of absorbing them with more efficiency.

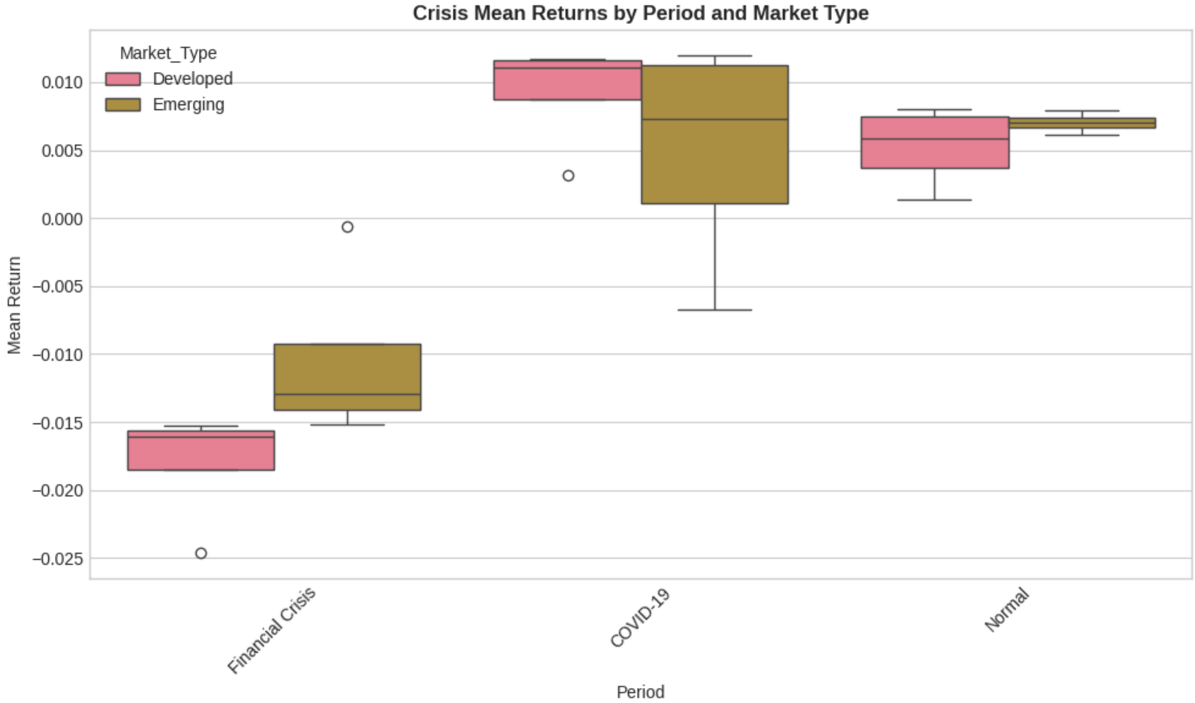

Fig. 9 Crisis vs. Normal Period Analysis: Crisis Mean Returns by Period and Market Type

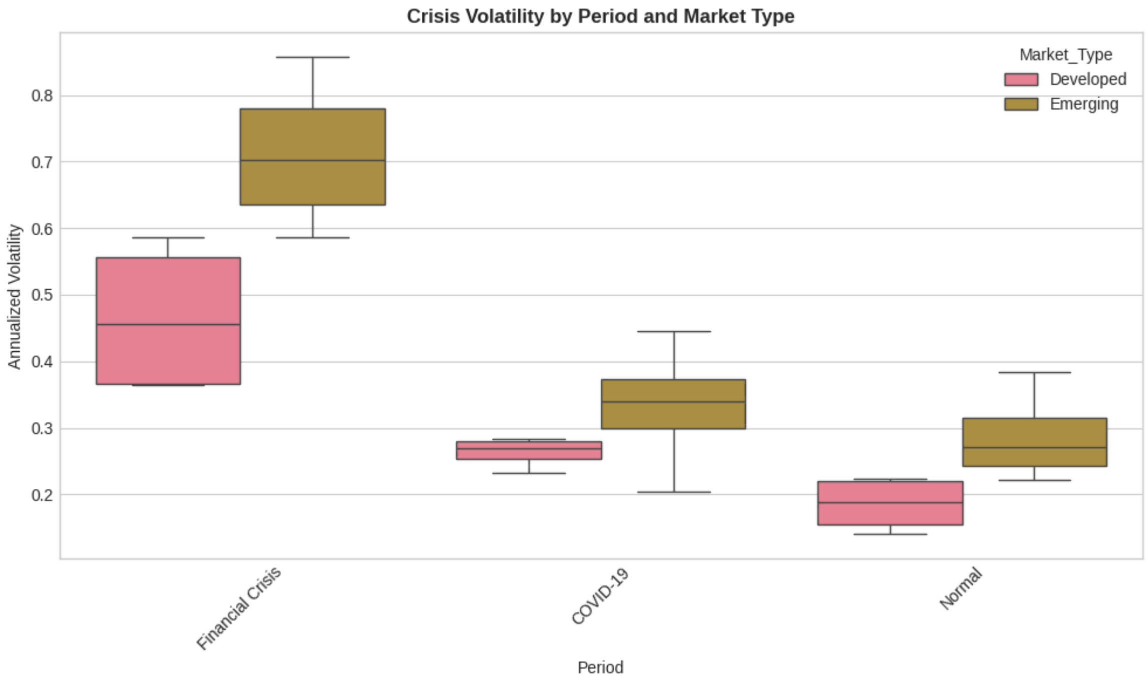

Fig. 10 Crisis vs. Normal Period Analysis: Crisis Volatility by Period and Market Type

8.4. Country-Level Comparison

Performance varies a lot across individual markets. In Fig. 11, the United States has an RMSE of 0.0567, while Brazil has 0.1107. In general, developed markets tend to have lower RMSE, whereas emerging markets are at the lower end. This further reinforces that developed financial systems are more stable in terms of predictability. Emerging markets are hindered by instability at this point in time.

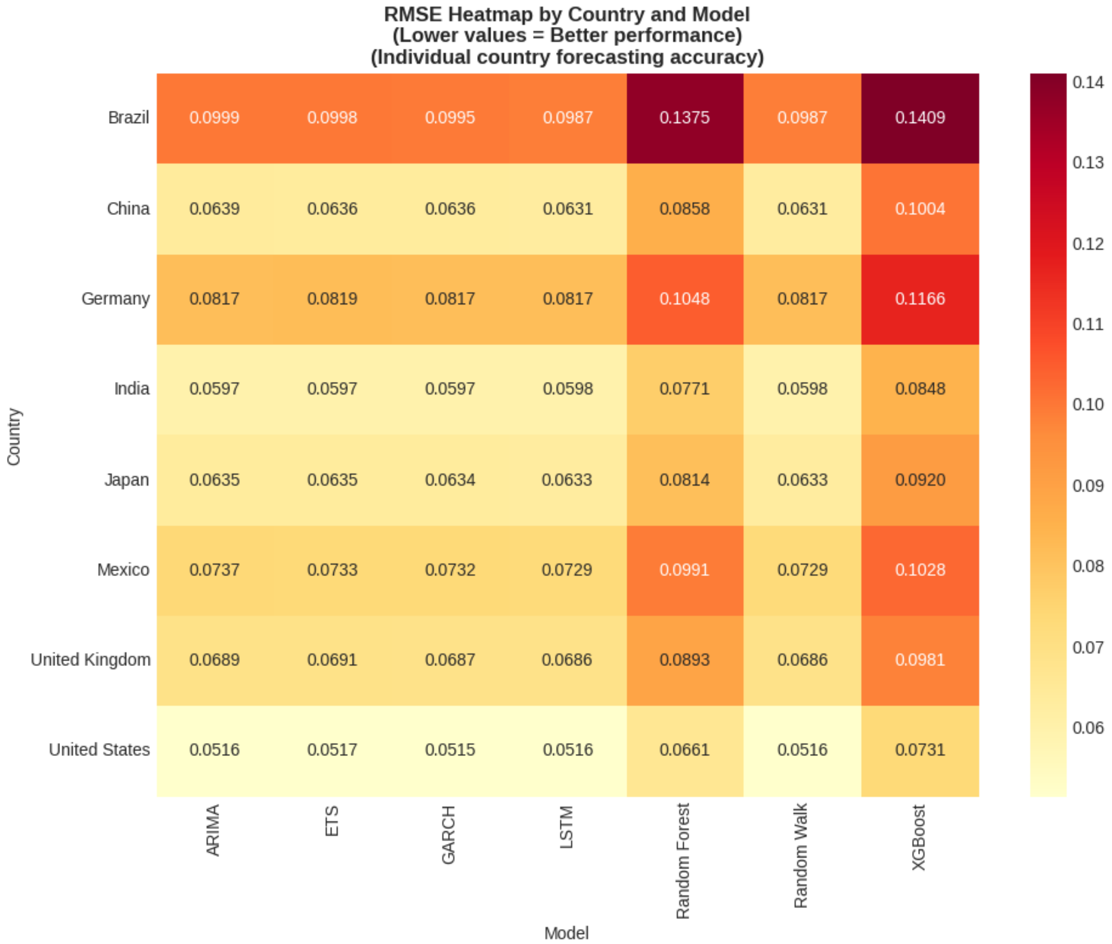

Fig. 11 RMSE Heatmap by Country and Model

8.5. Model Comparison

In all various models, shown in the bottom two plots in Fig. 12 and Fig. 13, the LSTM network consistently showed the lowest forecast errors in both emerging and developed markets (0.0736 and 0.0663, respectively). Econometric models such as ARIMA, ETS, and GARCH exhibited competitive performance in developed markets but were less effective in emerging markets. In comparison, tree-based machine learning models (Random Forest and XGBoost) performed relatively poorly in emerging markets. This indicates that emerging markets are more sensitive to noise and are overfitting in less stable systems. In addition, these results show that model choice interacts with market type and that deep learning is good at handling the nonlinearities in financial data.

Fig. 12 Model Performance Analysis: Developed Markets – Model Performance Spread

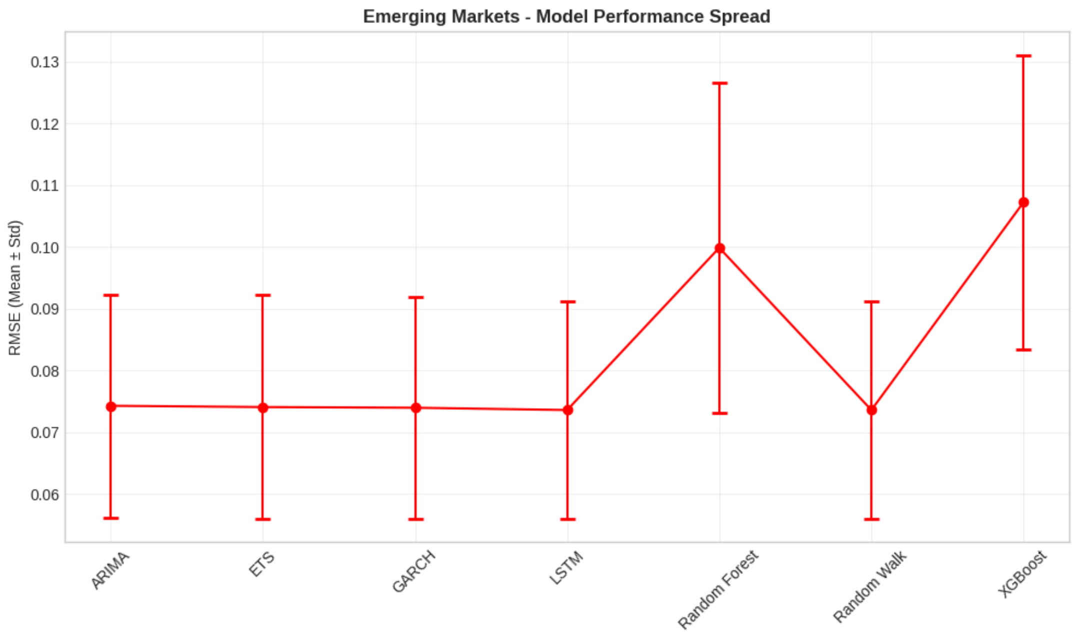

Fig. 13 Model Performance Analysis: Emerging Markets – Model Performance Spread

8.6. Economic Fundamentals and Predictability

In developed markets, the strongest predictors of stock returns were exchange rates, industrial production, and GDP. In emerging markets, by contrast, stock prices relative to the USD and real effective exchange rates (REER) were more predictive (Fig. 14). It is important to note that those economic fundamentals in emerging markets are more volatile than those in developed markets. The instability contributes to more forecast errors, since more volatile fundamentals diminish the signal-to-noise ratio for forecasting models.

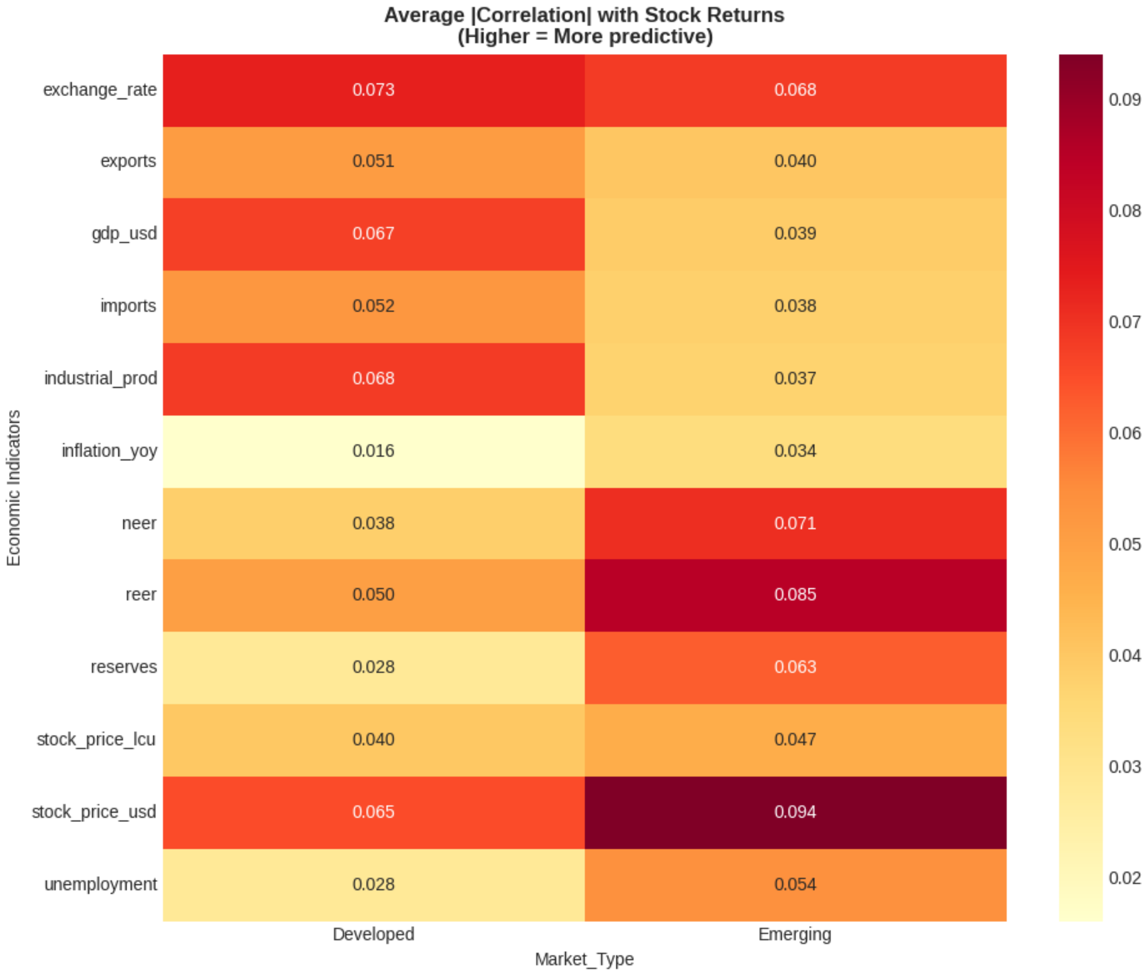

Fig. 14 Economic Indicator Analysis: Average Correlation with Stock Returns

Overall, the results contradict EMH predictions since emerging markets are not more predictable but are harder to forecast. On the other hand, efficiency indicators such as lower volatility and weaker autocorrelation show that developed markets are actually more efficient. This paradox suggests that market inefficiency does not necessarily mean higher predictability. Inefficiency is more likely to be related to volatility and instability that impact the accuracy of prediction.

9. Limitation and Future Directions

Despite the robustness of the design, several weaknesses remain. First, analysis is restricted to eight large economies, divided equally between advanced markets and emerging markets. Though these countries represent much of the world market capitalization, the comparatively small sample may limit generalizability, particularly for smaller or frontier markets. Second, the analysis is entirely conducted based on monthly data from the World Bank GEM dataset. While ensuring comparability by country, higher-frequency data (daily or intra-day) may reveal alternative dynamics, especially with the possibility of machine learning algorithms to reveal short-term nonlinear relationships. Third, implementations of models use fixed model specs (e.g., ARIMA(1,0,1), 50-tree Random Forest, baseline LSTM), and broader hyperparameter tuning could potentially alter the ranking of model performance. Fourth, results are analyzed by standard statistical error metrics such as RMSE, MAE, and MAPE. While suitable for benchmarking the accuracy of the model, they do not directly express economic significance by profitability, risk-adjusted return, or strategy profitability. Finally, while results for top-performing LSTMs were strongest for predictive outcomes, interpretation is restricted by these networks and thus questions the transparency of machine learning for financial applications.

Future research can take up many of these findings in various ways. Expanding coverage to additional world economies, including frontier markets, would subject the paradox of declining predictability of emerging markets to the test of scale. Use of higher-frequency financial data could be employed to capture short-horizon predictability and to bring the approach into finer correspondence with high-frequency trading. Methodologically, hybrid and ensemble models of the type that blend ARIMA-LSTM or GARCH-XGBoost with systematic hyperparameter tuning could boost predictive ability. Beyond accuracy, future research must attend to the economic value of the resulting forecasts through simulations of portfolios and risk-adjusted measures of results. More significantly, explainability techniques of the type that provide SHAP values or attention-based architecture could bridge the gap between predictive ability and model interpretability. Lastly, taking up time-varying efficiency representations would allow for a dynamic test of EMH and could shed light on how efficiency changes through regimes and through crises.

10. Conclusion

This research provides new evidence about stocks' unpredictability from mature and emerging economies. Burying Efficient Market Hypothesis hopes, evidence here indicates emerging economies' significantly greater forecasting errors and hence are neither informationally inefficient nor necessarily predictable. Instead, increased volatility and structural unpredictability seem to decrease forecast reliability to an extent comparable to state-of-the-art estimating techniques' findings. For all compared estimating models among all compared sets, LSTM networks always yielded superior performance for both sets of markets, and hence, there is value added by deep learning methods to reveal nonlinear financial data interdependencies. These are significant additions to the vast literature about market efficiency by providing evidence to shed light on whether and to what extent efficiency can mainly arise as volatility and less as exploitable regularities. From a practical point of view, implications are to cast doubt upon using the notion to treat emerging economies as necessarily and solely predictable and to shed light upon how much institutional maturity and research methodology can affect forecasting reliability.

References

[1] Bansal D. How AI is Transforming Stock Market Prediction: Key Insights [Internet]. Damco Solutions. 2023.

[2] Lin Y, Liu B. A Framework for Enhancing Stock Investment Performance by Predicting Important Trading Points with Return-Adaptive Piecewise Linear Representation and Batch Attention Multi-Scale Convolutional Recurrent Neural Network. Entropy. 2023, 25(11):1500–0.

[3] Risky Business. Why the Smart Money Forecasts Risk, Not Returns | Man Group [Internet]. Man.com. 2024.

[4] Fama E. Efficient Capital markets: a Review of Theory and Empirical Work. The Journal of Finance. 1970, 25(2):383–417.

[5] Chen J, Haboub A, Ali Shakil Khan. Limits of arbitrage and their impact on market efficiency: Evidence from China. Global Finance Journal. 2024, 1;59:100916–6.

[6] Rink K. The predictive ability of technical trading rules: an empirical analysis of developed and emerging equity markets. Financial Markets and Portfolio Management. 2023, 12;37(4):403–56.

[7] Bustos O, Pomares-Quimbaya A, Stellian R. Machine learning, stock market forecasting, and market efficiency: a comparative study. International Journal of Data Science and Analytics. 2025.

[8] Hanauer MX, Kalsbach T. Machine learning and the cross-section of emerging market stock returns. Emerging Markets Review. 2023, 101022.

[9] Orsel OE, Yamada SS. Comparative Study of Machine Learning Models for Stock Price Prediction [Internet]. arXiv.org. 2022.

[10] Htun HH, Biehl M, Petkov N. Survey of feature selection and extraction techniques for stock market prediction. Financial Innovation. 2023, 12;9(1).

[11] Li Z, Han J, Song Y. On the forecasting of high-frequency financial time series based on ARIMA model improved by deep learning. Journal of Forecasting. 2020, 39(7):1081–97.

[12] Xia F, Liu J, Nie H, Fu Y, Wan L, Kong X. Random Walks: A Review of Algorithms and Applications. IEEE Transactions on Emerging Topics in Computational Intelligence. 2020, 4(2):95–107.

[13] Kontopoulou VI, Panagopoulos AD, Kakkos I, Matsopoulos GK. A Review of ARIMA vs. Machine Learning Approaches for Time Series Forecasting in Data Driven Networks. Future Internet. 2023, 15(8):255.

[14] Benghiat S, Lahmiri S. Evaluating Predictive Models for Three Green Finance Markets: Insights from Statistical vs. Machine Learning Approaches. Computation. 2025, 13(3):76.

[15] Wang C, Gerlach R. A Bayesian realized threshold measurement GARCH framework for financial tail risk forecasting. Journal of Forecasting. 2023, 43(1):40–57.

[16] Breitung C. Automated stock picking using random forests. Journal of Empirical Finance [Internet]. 2023, 72:532–56.

[17] Ali AA, Khedr AM, El-Bannany M, Kanakkayil S. A Powerful Predicting Model for Financial Statement Fraud Based on Optimized XGBoost Ensemble Learning Technique. Applied Sciences [Internet]. 2023, 13(4):2272.

[18] Safwan Mahmood Al-Selwi, Mohd Fadzil Hassan, Said Jadid Abdulkadir, Amgad Muneer. LSTM Inefficiency in Long-Term Dependencies Regression Problems. Journal of Advanced Research in Applied Sciences and Engineering Technology. 2023, 30(3):16–31.

[19] Global Economic Monitor (GEM) | DataBank [Internet]. databank.worldbank.org. 2025.

[20] Almaiman MI. Forecasting Stock Prices Using ARIMA Models And Technical Analysis [Internet]. AFIT Scholar. 2024.